

# **Chapter 1**

## **Introduction**

This chapter reviews the literature before my thesis start date (11/2011), information published during the course of my thesis which was directly related to my thesis topic will be discussed as part of the Summary, conclusion and future direction in Chapter 7.

### **1.1 Greenhouse gases and global warming**

Greenhouse gases (GHGs) are atmospheric gases that absorb and emit radiation, warming the planet Earth and making it habitable to life (Karl and Trenberth 2003). GHGs include water vapour, carbon dioxide (CO<sub>2</sub>), ozone, methane (CH<sub>4</sub>) and other trace gases, which contribute 60%, 26%, 8% and 6% radiative forcing warming effect respectively (Kiehl and Trenberth 1997). Since the start of the industrial evolution in the 1760s, atmospheric concentrations of CO<sub>2</sub>, CH<sub>4</sub> and nitrous oxide have increased by 40%, 150% and 20% respectively and continue to rise today. Human activities involving fossil fuel usage, land use change and agriculture, are the main driver of the increasing GHGs (Cubasch *et al.* 2013). The averaged global combined land and ocean surface temperature showed an increase of 0.85°C between 1880 and 2012, consistent with the elevated GHG emissions due to anthropogenic activity (Hartmann *et al.* 2013). The elevated temperature directly contributes to the shrinking of glaciers, reduction of Arctic sea ice and ice sheets and the rise of global average sea level at an average rate of 1.8 mM per year (Church and White 2011). Global warming has also been linked to the increased frequency of hurricanes (Saunders and Lea 2008), as well as other extreme weather conditions (Rahmstorf and Coumou 2011; Schiermeier 2011). The elevated partial pressure of CO<sub>2</sub> has also caused ocean acidification (Fabry *et al.* 2008), as well as extinction of species and loss of biodiversity (Balint *et al.* 2011; Selbmann *et al.* 2012). Therefore, to limit the effects of global warming, it is necessary to reduce anthropogenic GHG emissions.

### **1.2 New Zealand agriculture and associated CH<sub>4</sub> emissions**

CH<sub>4</sub> is the second largest source of anthropogenic GHGs, with an atmospheric lifetime of 12.4 years and 28 times the global warming potential of CO<sub>2</sub> over 100 years (Cubasch *et al.* 2013). CH<sub>4</sub> emissions accounted for 44% of New Zealand's total GHG emissions in 2013, of which 90.4% came from agriculture (Ministry for the Environment [MfE], 2015). New Zealand is

unusual because it is one of the few developed country where agricultural GHG emissions play a major role in the national emissions profile. Agriculture was responsible for 32.5% of New Zealand's total exports in 2014 (www.stats.govt.nz 2015), which is sustained by a high number of livestock, including 29.8 million sheep, 6.7 million dairy cattle, 3.7 million beef cattle and 0.96 million deer (MacPherson 2014). The high number of ruminant livestock results in high production of CH<sub>4</sub> via enteric fermentation, accounting for 79.9% of New Zealand total CH<sub>4</sub> emissions (Ministry for the Environment [MfE], 2015). Because enteric CH<sub>4</sub> emissions account for such a large proportion of New Zealand's total GHG emissions, mitigation of enteric CH<sub>4</sub> emissions could have a significant effect in lowering the national GHG emissions. In order to formulate strategies for enteric CH<sub>4</sub> mitigation, it is necessary to understand the source of CH<sub>4</sub> and its formation in the rumen environment.

### **1.3 The rumen environment and CH<sub>4</sub> formation**

The stomach of ruminant animals, such as cattle, sheep, deer and goats, is divided into four compartments that perform different digestive functions. The largest compartment, called the rumen, is where symbiotic microorganisms breakdown the digesta; the reticulum controls the movement of digesta as well as containing a small amount of digesta and microorganisms, the omasum absorbs nutrients and re-absorbs water, and the abomasum is the true acid stomach that secretes acidic digestive juices (Boomker and Cronjé 2000). The digestion of feed is carried out mainly in the rumen and reticulum by a combination of rumination and enzymatic degradation by rumen microorganisms. Ruminants do not possess the enzymes necessary to digest the structural carbohydrate components of plant cell walls, such as cellulose, hemicelluloses and pectin. Ruminants rely on the complex microbial community within their rumen to hydrolyze and ferment plant polysaccharides (Hungate 1966). The microbial community in the rumen and reticulum includes bacteria, archaea fungi, viruses, and protozoa, the number and diversity of which is as rich as any other natural microbial habitat (Hungate 1966). Ruminant animals and their rumen microbes have achieved a mutually beneficial symbiosis, in which the animal benefits from the volatile fatty acids (VFAs), acetate, propionate and butyrate which are the products of forage digestion by microbes, while the microbes benefit from a stable, buffered environment in which they are able to grow (Hungate 1966).

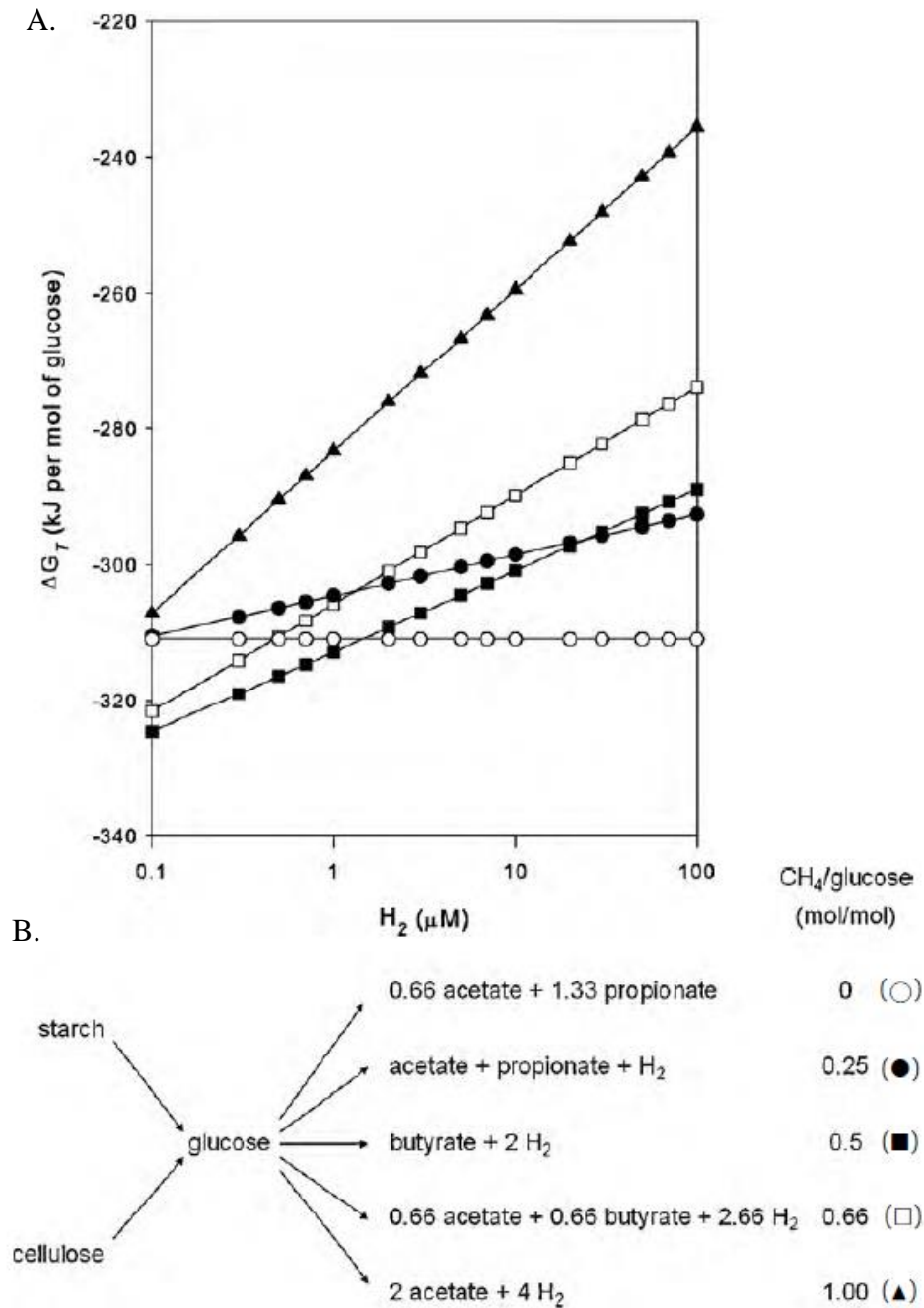
In addition to the VFAs mentioned above, there are also intermediate fermentation products that are produced and consumed continuously, such as succinate, lactate, formate, ethanol,

methanol, methylamines, and hydrogen ( $H_2$ ). Several of these intermediary compounds are used by rumen methanogenic archaea to produce  $CH_4$  (Hungate 1967).

The abundance of rumen archaea within the ovine, bovine and caprine rumen ranges from 0.3% to 3.3% of total rumen microorganisms, with an average of 1.69%, as estimated by archaea-specific DNA probes targeting a conserved region of the small subunit ribosomal RNA genes (Lin *et al.* 1997). It has also been observed that 2.8% to 4.0% of rumen microorganisms showed  $F_{420}$  autofluorescence, a unique property of methanogens (Yanagita *et al.* 2000), indicating that the majority of rumen archaea appear to be methanogens.

Methanogenic archaea depend solely on methanogenesis for energy generation, and can be classified into three categories based on the mechanism used. i) Hydrogenotrophic methanogens are predominant in the rumen where they use  $H_2$  (or formate) as an energy source and reduce  $CO_2$  to  $CH_4$ . This use of  $H_2$  prevents its accumulation and an increase of  $H_2$  partial pressure that can inhibit microbial fermentation (Hungate 1967). ii) Methylotrophic methanogens use methylated compounds such as methanol, methylamine and methylthiol as substrates for  $CH_4$  formation. iii) The acetoclastic methanogens are the least common in the rumen, and they use acetate for methanogenesis which is energetically inefficient for the animal (Thauer 1998). It is believed that methanogenic archaea that depend solely on acetoclastic methanogenesis likely grow too slow to persist under normal conditions in the rumen and disappear from the rumen due to the wash out effect (Rowe *et al.* 1979; Hobson and Stewart 1997).

A model of fermentation has been proposed in which the hydrogen concentration has a strong influence on the reactions that are thermodynamically favoured within the rumen (Janssen 2010). Figure 1.1 shows the main pathways for VFA and  $H_2$  production and demonstrates how these reactions are influenced by changing ruminal  $H_2$  concentrations.



**Figure 1.1 The thermodynamic relationship between hydrogen concentration and VFA production from cellulose and starch.** A. Gibbs free energy changes ( $\Delta G_T$ ) of glucose fermentation at different  $H_2$  concentrations. B. Possible pathway of cellulose and starch fermentation via glucose to acetate, propionate and butyrate, for simplicity,  $CO_2$ ,  $H^+$  and  $H_2O$  are not shown (Janssen 2010).



Because most rumen methanogens use  $H_2$ , methanogen activity and growth are intricately linked to hydrogen producing microbes, such as rumen protozoa and bacteria, whose activity is influenced by the feed eaten by the animal. In New Zealand, ruminants are fed mainly pasture forages, while conserved forms of pasture forages (silage or hay) or forage crops (maize, brassicas) are also important at various times of the year. Less commonly, grains or mixtures of the previous feeds, are fed. Pasture forages typically contains more cellulose and hemicellulose, compared to grains, and its slower rate of fermentation in the rumen results in more  $CH_4$  formation. Grains are rich in starch and are broken down more readily by rumen microorganisms, producing more propionate when fermented in the rumen (Huntington 1997), acting as an alternative sink for electrons rather than going to  $CH_4$  (Moss *et al.* 2000). The general observation is that the slower the rate of digestion in the rumen, the more  $CH_4$  is formed (Moe and Tyrrell 1979), and that forage has a slower passage rate through the rumen, and the high cellulose content favours the formation of acetate (Janssen 2010).

Understanding the characteristics of rumen methanogens and their complex relationship with other rumen microbes, are therefore essential to formulate strategies to mitigate enteric  $CH_4$  emissions, either directly against methanogens or indirectly against rumen microbes supporting methanogen activities.

## **1.4 $CH_4$ mitigation strategies**

$CH_4$  mitigation strategies must consider the entire rumen microbial community and its contribution to the nutrition of the host. Strategies must be viable without reducing animal production or have a negative impact on the farming operation (Clark *et al.* 2011). The products of animals (most commonly milk, meat and wool) must not contain residues of any compound used to reduce  $CH_4$  emissions. It is likely that methanogens species in the rumen occupy different niches and are adapted to grow at different levels of  $H_2$  and/or on different substrates, therefore mitigation strategies must be able to tackle methanogens living under different circumstances. This applies to their physical location as some methanogens are free-living while others are attached to, or live within hydrogen producing protozoa (Fenchel and Finlay 1991). Finally, if methanogens are eliminated from the rumen, then an alternative  $H_2$  disposal method must take the place of the methanogenesis, otherwise the ruminant host can suffer from accumulation of  $H_2$ , reduced fibre degradation and lower fermentation rates.

There are many CH<sub>4</sub> mitigation strategies that indirectly reduce methanogens by reducing H<sub>2</sub> production, such as redirecting VFA production from acetate to propionate (McAllister and Newbold 2008), or removal of hydrogen producing protozoa and cellulolytic bacteria (Morgavi *et al.* 2012). Low fibre, high starch grain diets such as barley, result in faster digestion time and reduced energy loss via CH<sub>4</sub> from 6.5% to 3% (Beauchemin and McGinn 2005). While this feeding strategy is cost effective in countries with abundant supply of cheap grains, it is not applicable in New Zealand which depends heavily on forage grazing. Feed additive such as canola oil, coconut oil, sunflower oil have also been shown to reduce CH<sub>4</sub> production (McGinn *et al.* 2004), as fatty acids may act as alternative hydrogen sinks (Johnson and Johnson 1995), however this method is not ideal as it lowers the animal performance by decreasing fibre digestibility (Beauchemin and McGinn 2006). Application of essential oils as feed additives have reported similar findings (Patra and Yu 2012).

Many feed additives have been tested for CH<sub>4</sub> mitigation. Antibiotic ionophores such as monensin, inhibit protozoa and cellulolytic bacteria (Hino *et al.* 1993), therefore the introduction of monensin as a feed additive is thought to be able to reduce H<sub>2</sub> production and as a result mitigate CH<sub>4</sub> production by up to 10% (Beauchemin *et al.* 2008). However, usage of antibiotics is banned by the European Union, because of the risk that antibiotics could get absorbed from the rumen and contaminate milk and meat (Pugh 2002). Furthermore, the reductions in CH<sub>4</sub> have been observed to be transient (Guan *et al.* 2006), and prolonged use of antibiotics could potentially contribute to antibiotic resistance. Organic acids, such as fumarate and malate, can be used as feed additives to redirect bacterial VFA production from acetate to propionate, as many rumen bacteria are known to produce propionate via a reverse citric acid cycle, of which malate and fumarate are key intermediates in the pathway (Martin 1998). Despite the reduction of CH<sub>4</sub> production by fumarate and malate observed *in vitro* (Newbold *et al.* 2005; Lin *et al.* 2013), the *in vivo* trials thus far remain inconclusive (Beauchemin and McGinn 2006; Foley *et al.* 2009). Cashew nut shell liquid has also been tested as a feed additive as it contains high amounts of anacardic acid, which inhibits Gram-positive bacteria (Kubo *et al.* 2003), redirecting rumen fermentation to propionate production and reducing CH<sub>4</sub> production, although reduced digestibility has been observed (Shinkai *et al.* 2012).

Protozoa are one of the primary H<sub>2</sub> producers in the rumen, and methanogens have been shown to be endo- and ectosymbiotic to protozoa (Finlay *et al.* 1994). Endosymbiotic methanogens within protozoa could account for 10% to 20% of rumen methanogens (Chagan *et al.* 1999). Therefore removal of protozoa could be a viable strategy for CH<sub>4</sub> mitigation. The removal of

protozoa from the rumen is known as defaunation and several different methods exist, including chemical drenching and diet alteration to lower the pH in the rumen. The *in vivo* results for reducing CH<sub>4</sub> via defaunation thus far are inconsistent and it is debatable whether defaunation is capable of reducing CH<sub>4</sub> emissions (Bird *et al.* 2008). Defaunation is likely unsuitable in New Zealand as a CH<sub>4</sub> mitigation option, because pasture-fed animals tend to have lower number of protozoa (Towne *et al.* 1990; Silva *et al.* 2014), and it would be very difficult to prevent re-faunation of the rumen from natural sources.

Another method to reduce protozoa is the incorporation of plant secondary metabolites, such as saponins and tannins, as feed additives. Saponins are natural detergents with emulsification properties and are abundant in tea leaves. Saponins significantly reduce protozoa populations (Guo *et al.* 2008) and possibly impair protozoal membranes by forming complexes with sterols (Wallace *et al.* 2002). Overall, the effect of saponin addition on CH<sub>4</sub> reduction has been inconsistent (Jayanegara *et al.* 2014). Tannins are a class of polyphenols of high molecular weight, and are abundant in tropical legumes. The effect of tannins on protozoa and methanogens is inconsistent. Sheep feed on plants containing high levels of tannins have up to 30% reduction in CH<sub>4</sub> production without adverse effects (Carulla *et al.* 2005; Tavendale *et al.* 2005), and both methanogens and protozoa significantly declined as the concentration of tannin was increased in the feed (Tan *et al.* 2011). However, some studies also found tannins to have no effect on CH<sub>4</sub> reduction (Beauchemin *et al.* 2007). Therefore increasing the proportion of tannin-rich plants in forage is applicable in New Zealand, but it would require more research with statistical support to ascertain its viability as a CH<sub>4</sub> mitigation option.

The traditional method of selective animal breeding could also reduce CH<sub>4</sub> emission, if CH<sub>4</sub> emissions are a heritable trait. Selecting for feed conversion efficiency and productivity may also reduce CH<sub>4</sub> emissions per unit of product formed. Changes in farming practices, such as extending lactation in the dairy industry, could also reduce the energy demand from herd and reduce overall emissions (Eckard *et al.* 2010; Buddle *et al.* 2011).

The anaerobic cultivation technique developed by Robert Hungate in 1947, allowed the cultivation of rumen methanogens *in vitro*, yet it remains difficult to cultivate methanogens *in vitro* even today. This has stunted the development of CH<sub>4</sub> mitigation strategies directly targeting rumen methanogens. The advancement in sequencing technologies has enabled culture-independent methods to evolve, and has made CH<sub>4</sub> mitigation strategies directly targeting rumen methanogens possible. One example is the use of methanogen genomic

information to develop vaccines that specifically target rumen methanogens (Attwood *et al.* 2008; Buddle *et al.* 2011). This approach relies on the use of specific methanogen proteins to elicit a secretory antibody response in the saliva of ruminants. Cattle secretes approximately 1.5 to 2.5 rumen volumes of saliva per day (Bailey and Balch 1961), thus a vaccinated animal would secrete antibodies against methanogens via its saliva. This would achieve a sustained inhibition and prevent re-colonization of rumen methanogens (Subharat *et al.* 2015), providing a practical approach for grazing ruminants in New Zealand. Another example is the development of small molecule inhibitors, via a chemogenomics approach, targeting pathways and enzymes unique to rumen methanogens. However this approach needs to overcome the dilution effect of rumen as well as requirement to continuously deliver the inhibitor into the animal to prevent re-colonization of rumen methanogens (Leahy *et al.* 2010; Aung *et al.* 2015).

The ideal targets for vaccination and inhibitors are genes conserved across all rumen methanogens which are absent from the animal host and other members of the rumen microbiome (Attwood *et al.* 2008). All rumen methanogens need to be targeted because if any are not inhibited, they would be able to quickly adapt and populate the vacant niche. One possible target is the methyl CoM reductase gene, which is inhibited by the coenzyme M (CoM) analogue inhibitor, 2-bromoethanesulfonic acid (BES) (Immig *et al.* 1996). The genome sequences of rumen methanogens can improve our understanding of their biology and identify conserved genes that may serve as targets for developing effective CH<sub>4</sub> mitigation technologies.

Methanogens are not the only hydrogen utilisers in the rumen, and in the absence of methanogens the rumen microbiome would require an alternative hydrogen sink so the host digestion would not be impaired. Homoacetogens are organisms that consume H<sub>2</sub> and produce acetate, acting as alternative hydrogen sink and increasing the available energy to the host by 4% to 12% (Joblin 1999). However, the H<sub>2</sub> level in the rumen strongly favours methanogenesis over acetogenesis due to the H<sub>2</sub> threshold and the energy generated. The energy yield from methanogenesis is four times higher than acetogenesis (Thauer 1998; Muller 2003), therefore in the natural rumen environment, homoacetogens cannot compete with methanogens for H<sub>2</sub>. Despite the abundance of strategies in methane mitigation, our lack of understanding of the rumen microbiome limits the capability to develop effective long-term strategies.

## 1.5 Rumen methanogens

In order to formulate effective strategies of CH<sub>4</sub> mitigation, it is necessary to understand the broadness of diversity of rumen methanogens as well as their variety of metabolic processes. Methanogenic archaea belong to the Euryarchaeota phylum, which consists of seven orders; the Methanobacteriales, Methanomicrobiales, Methanosarcinales and Methanomassiliicoccales are found in the rumen as well as other environments, while the Methanocellales, Methanococcales and Methanopyrales are found outside of the rumen (Liu and Whitman 2008; Sakai *et al.* 2008; Iino *et al.* 2013). There are at least 28 genera and 113 species of methanogens (Garriety *et al.* 2007).

A survey of published 16S rRNA gene sequences has shown that rumen methanogenic archaea form nine separate clades (Janssen and Kirs 2008). Globally, the majority (92.3%) of rumen archaeal sequences fall within three main methanogen groups, *Methanobrevibacter*, *Methanomicrobium* and of the newly established Methanomassiliicoccales order. *Methanobrevibacter* is the predominant genus of rumen methanogens, accounting for 61.6% of rumen archaea. Around 33.6% of archaeal 16S rRNA gene sequences fall within the *Mbb. gottschalkii* clade and 27.3% in the *Mbb. ruminantium* clade (Janssen and Kirs 2008). Members of Methanomassiliicoccales order account for 15.8% of global rumen archaeal population, and can make up to 85% of rumen methanogens in particular ruminants (Huang *et al.* 2012), while *Methanomicrobium* account for 14.9% of the rumen archaeal population. The remaining low abundance groups in the rumen includes *Methanomicrococcus* spp., *Methanobacterium* spp., and *Methanosarcinales* spp. (Yanagita *et al.* 2000; Shin *et al.* 2004; Wright *et al.* 2007; Janssen and Kirs 2008). In New Zealand ruminants, the Methanobacteriales account for 89.6% of rumen methanogens, of which the *Mbb. gottschalkii* clade accounts for 42.4% and *Mbb. ruminantium* clade accounts for 32.9% (Seedorf *et al.* 2015). Methanomassiliicoccales account for 10.4% of methanogens in New Zealand ruminants and is the second most abundant order (Seedorf *et al.* 2015).

Only a handful of rumen methanogens have been isolated and grown in pure cultures due to difficulties in culturing these strict anaerobes. The majority of cultures belong to the predominant genus, *Methanobrevibacter*, and display morphologies of short rods with a tendency to form pairs, chains or irregular clumps (Fournier *et al.* 2011). The first rumen methanogen described was *Methanobacterium ruminantium*, and was isolated in 1958 using serial dilution in a rumen simulating medium (Smith and Hungate 1958). The original strain

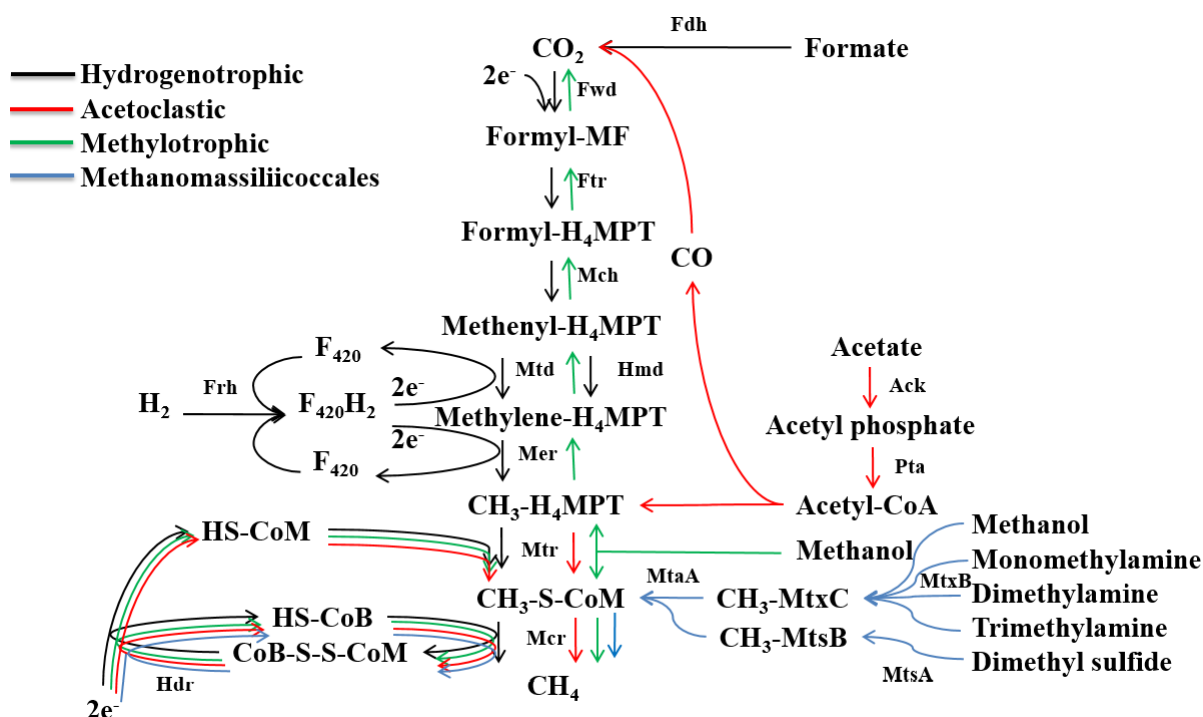
was lost, but was re-isolated in 1965 (Bryant 1965), and subsequent verification and phylogenomic study renamed this species as *Mbb. ruminantium* M1<sup>T</sup> (Balch *et al.* 1979). Subsequently, several species and strains belonging to the *Mbb.* genus have been isolated from a variety of ruminant sources. A member of the *Mbb. wolinii* clade, *Mbb.* sp. AbM4, has been isolated from ovine abomasum (Simcock *et al.* 1999). Four formate utilizing methanogens were isolated in 2007, all belonging to the genus, *Methanobrevibacter* (Rea *et al.* 2007). Three isolates were described as strains of *Mbb. olleyae*, two of which were from a bovine rumen source and one from the ovine rumen. The fourth isolate was described as *Mbb. millerae* from the bovine rumen. A member of the *Mbb. ruminantium* clade, *Mbb. olleyae* YLM1, and representatives of the *Mbb. gottschalkii* clade, *Mbb.* sp. D5, and *Mbb.* sp. SM9 have been isolated from ovine rumen (Kim 2012). Three strains of *Mbb. boviskoreani* have been isolated from bovine rumen in South Korea (Lee *et al.* 2013).

Other groups of methanogens, such as *Methanomicrococcus*, *Methanosphaera* and *Methanobacterium* were either less abundant or were not detected in all studies (Yanagita *et al.* 2000; Shin *et al.* 2004; Wright *et al.* 2007; Janssen and Kirs 2008). The family *Methanobacteriaceae*, which includes the genera *Methanobrevibacter*, *Methanobacterium*, and *Methanosphaera*, composed from 30% to 99% of rumen archaea, while the family *Methanomicrobiales* account for 0 to 54% of rumen archaea. *Methanobacterium formicium*, *Methanomicrobium mobile* have been isolated as pure cultures from the bovine rumen (Jarvis *et al.* 2000). The family *Methanosarcinales* were a small minority that ranged from 2% to 3% of rumen archaea (Janssen and Kirs 2008), similar conclusions have been drawn from FISH studies (Soliva *et al.* 2003), *Methanosarcina barkeri* have been isolated from the bovine rumen as pure cultures (Jarvis *et al.* 2000). *Methanosphaera* is present in the rumen of deer, sheep and cattle, and studies have found there are two major clusters of *Methanosphaera* spp., a particular strain ISO3-F5 has been isolated from ovine rumen and is cultivated *in vitro* (Jeyanathan *et al.* 2011).

CH<sub>4</sub> mitigation strategies often involves *in vitro* experimental validation on methanogen cultures, such as testing potential methanogen-inhibitory compounds, therefore the scarcity of pure cultures of rumen methanogens is a limitation, making it difficult to validate whether the strategies could effectively inhibit all methanogens within the rumen, which also highlights the importance of DNA sequencing efforts in the formulation of CH<sub>4</sub> mitigation strategies.

## 1.6 Methanogenesis

Methanogenesis is the sole energy generating mechanism of rumen methanogens, therefore conserved enzymes in the methanogenesis pathways would make suitable targets for methanogen inhibitors. The pathways of hydrogenotrophic, acetoclastic, methylotrophic methanogenesis as well as the methanogenesis pathway of *Methanomassiliicoccales* are displayed in Figure 1.2 (Thauer 1998; Borrel *et al.* 2014).



**Figure 1.2 Methanogenesis pathways.** NAD-dependent formate dehydrogenase (Fdh). Formylmethanofuran dehydrogenase (Fwd). Formylmethanofuran (MF). Formylmethanofuran:H<sub>4</sub>MPT formyltransferase (Ftr). Tetrahydromethanopterin (H<sub>4</sub>MPT). Methenyl- H<sub>4</sub>MPT cyclohydrolase (Mch). Coenzyme F<sub>420</sub>, 8-hydroxy-f-deazaflavin (F<sub>420</sub>). Coenzyme F<sub>420</sub>-reducing hydrogenase (Frh). F<sub>420</sub>-dependent methylene-H<sub>4</sub>MPT dehydrogenase (Mtd). H<sub>2</sub>-forming methylene-H<sub>4</sub>MPT dehydrogenase (Hmd). F<sub>420</sub>-dependent methylene-H<sub>4</sub>MPT reductase (Mer). Coenzyme M, 2-sulfanylethanesulfonate (CoM). Methyl-H<sub>4</sub>MPT: coenzyme M methyltransferase (Mtr). Methyl:coenzyme M reductase (Mcr or Mrt). Acetate kinase (Ack). Phosphoacetyl transferase (Pta). Methanol/mono-, di-, tri-methylamine corrinoid methyltransferase (MtxB). Methanol/mono-, di-, tri-methylamine corrinoid protein (MtxC). Methylthiol:corrinoid methyltransferase (MtsA). Methylthiol:corrinoid protein (MtsB). Methyl:coenzymeM methyltransferase (MtaA). Heterodisulfide reductase (Hdr). Coenzyme B, 7-mercaptoheptanoylthreoninephosphate (CoB) (Thauer 1998).

Hydrogenotrophic methanogenesis is the most common methanogenic pathway observed in rumen methanogens, and is utilised by the predominant *Methanobrevibacter* spp. (Leahy *et al.* 2010; Leahy *et al.* 2013). In the hydrogenotrophic pathway, CO<sub>2</sub> is reduced to CH<sub>4</sub> through a series of enzymatic steps as illustrated in Figure 1.2. The initial reduction of CO<sub>2</sub> with the C<sub>1</sub> carrier methanofuran to formyl methanofuran by formylmethanofuran dehydrogenase (Fwd) is driven by a sodium gradient in *Methanothermobacter thermoautotrophicus* (de Poorter *et al.*

2003). The tetrahydromethanopterin (H<sub>4</sub>MPT) is the second C<sub>1</sub> carrier, and the transfer of the formyl group is catalysed by formylmethanofuran:H<sub>4</sub>MPT formyltransferase (Ftr) (Shima *et al.* 2002). The formyl-H<sub>4</sub>MPT is then reduced to methyl-H<sub>4</sub>MPT via methenyl-H<sub>4</sub>MPT cyclohydrolase (Mch), F<sub>420</sub>-dependent methylene-H<sub>4</sub>MPT dehydrogenase (Mtd), and F<sub>420</sub>-dependent methylene-H<sub>4</sub>MPT reductase (Mer) (Shima *et al.* 2002). The Mtd and Mer depend on reducing potential supplied from H<sub>2</sub> via F<sub>420</sub>-reducing hydrogenase (Frh). The Frh depends on the cofactor F<sub>420</sub> (Cheeseman *et al.* 1972), which is so named due to its autofluorescence at 420 nm. This property is particularly useful in enabling visualisation of rumen methanogens under UV illumination (Doddema and Vogels 1978). The methyl group of methyl-H<sub>4</sub>MPT is transferred to CoM via the transmembrane methyl-H<sub>4</sub>MPT: coenzyme M methyltransferase (Mtr), which drives a sodium gradient (Gottschalk and Thauer 2001). The membrane potential generated from Fwd and Mtr contribute to the membrane potential involved in energy conservation. The formation of methyl-CoM, the recycling of CoM and the formation of CH<sub>4</sub> is conserved with the other methanogenesis pathways. In the acetoclastic pathway, acetate is first converted to acetyl-CoA by acetyl-CoA synthetase or by acetate kinase (Ack) and phosphoacetyl transferase (Pta). Carbon monoxide dehydrogenase then cleaves acetyl-CoA, releases CO and transfers the methyl group to H<sub>4</sub>MPT, entering the methanogenesis pathway at Mtr. The CO is oxidized by ferredoxin and hydrogen, producing CO<sub>2</sub>, which feeds into the methanogenesis pathway (Fournier and Gogarten 2008). In methylotrophic methanogenesis, the methyl group from different C<sub>1</sub> substrates is transferred to CoM and enters the methanogenesis pathway at Mtr. Members of the orders Methanosarcinales and Methanomassiliicoccales and the genus *Methanosphaera* are capable of carrying out methylotrophic methanogenesis. The Methanosarcinales is the only order capable of disproportionating methanol to CH<sub>4</sub> and CO<sub>2</sub> as depicted in Figure 1.2 (Keltjens and Vogels 1993). Members of *Methanosphaera* are dependent on methylotrophic methanogenesis and are incapable of disproportionating methanol to CO<sub>2</sub> or utilising CO<sub>2</sub> for methanogenesis. The genome sequence of *Methanosphaera stadtmanae* confirmed this is due to absence of genes encoding the molybdenum cofactor biosynthesis complex, which is important in producing the essential cofactor to the formylmethanofuran dehydrogenase in the hydrogenotrophic methanogenesis pathway (Fricke *et al.* 2006). Members of the order Methanomassiliicoccales also depend solely on methylotrophic methanogenesis for survival, but differ from the *Methanosphaera* in that they do not possess any of the genes encoding the hydrogenotrophic part of the methanogenesis pathway (Borrel *et al.* 2014) (Figure 1.2). Therefore the Mtr/Mcr is the only enzyme shared by all the known methanogenesis pathways in rumen methanogens. Inhibitors affecting this



enzyme have been tested, including the CoM analogue BES and the CH<sub>4</sub> analogue, chloroform. BES inhibition of methanogens in the rumen is temporary, followed by quick adaptation of the methanogen community (Immig *et al.* 1996), while chloroform is toxic to the host animal (Lanigan *et al.* 1978).

In the methyltrophic methanogenesis pathway, the utilisation of each substrate involves three proteins carrying out transfer of methyl groups from substrates to the CoM. These include the substrate:corrinoid methyltransferases MtaB, MtmB, MtbB, MttB, MtsA for methanol, mono-, di-, tri- methylamine, methylthiol respectively, and the corrinoid proteins MtaC, MtmC, MtbC, MttC, MtsB for methanol, mono-, di-, tri- methylamine, methylthiol respectively (Burke and Krzycki 1997; Ferguson and Krzycki 1997; Sauer and Thauer 1998; Ferguson *et al.* 2000; Bose *et al.* 2008). The methanol:coenzyme M methyltransferase, MtaA, is shared by different substrates, except for methylthiol which uses the MtsA, a bifunctional enzyme which carries out both the transfer of methyl group from substrate onto corrinoid protein as well as the transfer from the corrinoid protein onto CoM (Tallant and Krzycki 1997; Tallant *et al.* 2001). The corrinoid proteins depend on the cobalt ion in their active sites to be in a reduced state, and a methyltransferase-activating protein, RamA, is responsible for activation of the corrinoid protein (Wassenaar *et al.* 1996). The remainder of the methylotrophic pathway is conserved with the hydrogenotrophic and acetoclastic methanogenesis pathways.

The formation of methyl-CoM is the point of convergence of methanogenesis pathways. The methyl-coenzyme M reductase is responsible for the final reduction of the methyl group to CH<sub>4</sub>. The functional components was first elucidated from *Methanobacterium thermoautotrophicum* (Gunsalus and Wolfe 1980). Two isozymes of this enzyme have been observed in *M. thermoautotrophicum*  $\Delta$ H (Rospert *et al.* 1990), Mcr (MCR I) and Mrt (MCR II). Both isozymes are composed of  $\alpha$ ,  $\beta$  and  $\gamma$  subunits encoded by *mcrA*, *mcrB* and *mcrG* respectively, yet the isozymes differ in their operon arrangements, *mcrBDCGA* and *mrtBDGA*. The isozymes are differentially regulated by growth condition, with Mcr up-regulated during growth on limiting H<sub>2</sub>, and Mrt up-regulated in abundant H<sub>2</sub> (Pihl *et al.* 1994; Reeve *et al.* 1997).

Not all members of Methanobacteriales possess both isozymes; *Mbb. ruminantium* M1<sup>T</sup> possess only the Mcr, which suggested it has adapted to a lifestyle at low H<sub>2</sub> availability (Leahy *et al.* 2010). The functional methyl-CoM reductase enzyme is a hexamer consisting of subunits Mcr/MrtA<sub>2</sub>B<sub>2</sub>G<sub>2</sub> with nickel containing cofactor F<sub>430</sub> in the active site. The function of the

McrCD and MrtD subunits remain elusive, but it has been proposed that their role may be in post-translational modification of Mcr/MrtA (Ellefson *et al.* 1982; Ermler *et al.* 1997).

The reduction of methyl-CoM to CH<sub>4</sub> is coupled to the formation of the heterodisulfide, CoM-S-S-CoB, which is then reduced by the heterodisulfide reductase complex (Hdr) to produce CoM-SH and allow it to be reused in methanogenesis (Noll *et al.* 1986; Hedderich *et al.* 1990; Hedderich *et al.* 2005). Two components of Hdr exist, the cytoplasmic HdrABC and the transmembrane HdrDE, which was found in *Methanosarcina barkeri*, with the HdrE being identified as a *b*-type cytochrome (Heiden *et al.* 1994).

The reducing equivalents required for heterodisulfide reduction come from the H<sub>2</sub>. HdrABC gains reducing potential either from Frh or from methylviologen hydrogenase (Mvh). As indicated above, Frh supplies reducing potential from H<sub>2</sub> by reducing F<sub>420</sub>. The Mvh complex, MvhADG, does not require cofactor F<sub>420</sub> to function. HdrABC has been observed to form a complex with MvhADG in *Methanobacterium thermoautotrophicum* (strain Marburg), and the complex was found to be loosely membrane bound and proposed to be associated with an integral membrane proton pump (Fiebig and Friedrich 1989; Setzke *et al.* 1994). In *Methanosarcina barkeri*, the transmembrane HdrDE has been found as a complex with membrane associated energy-converting hydrogenase (EchABCDE), which reduces ferredoxin with H<sub>2</sub> (Kunkel *et al.* 1998; Hedderich 2004). The periplasmic methanophenazine-dependent hydrogenase, Vht and Vho, has also been found in *Methanosarcina mazei* in addition to the Ech complex, and was found to be differentially expressed dependent on utilisation of acetate (Guss *et al.* 2009).

The hydrogenases not only supply reducing potential to methanogenesis, they are also an important feature linking methanogenesis to energy conservation. The F<sub>420</sub>H<sub>2</sub>, reduced ferredoxins or cytochromes produced by hydrogenases act in conjunction with proton pumps to link methanogenesis with membrane potential generation, which enables ATP synthase to generate energy (Deppenmeier *et al.* 1999; Deppenmeier 2002). The F<sub>420</sub>H<sub>2</sub> dehydrogenase complex (Fpo) is a large transmembrane proton pump, and in *Methanosarcina mazei* it is a key enzyme that generates membrane potential by electron transport from F<sub>420</sub>H<sub>2</sub> (Baumer *et al.* 2000; Welte and Deppenmeier 2011), as well as being linked to Hdr by electron transporting methanophenazine (Tietze *et al.* 2003). The membrane potential required for energy generation can be Na<sup>+</sup> or H<sup>+</sup>. In the hydrogenotrophic methanogens, the Mtr complex generates Na<sup>+</sup> gradient (Lienard *et al.* 1996). The Fwd is driven by the Na<sup>+</sup> gradient in hydrogenotrophic

methanogenesis (Kaesler and Schonheit 1989; de Poorter *et al.* 2003), and may act in reverse to generate  $\text{Na}^+$  gradient when the methyl substrate is disproportionated in methylotrophic methanogenesis (Deppenmeier *et al.* 1999). The  $\text{H}^+$  gradient is generated by the Fpo complex powered by the Hdr and Mvh (Deppenmeier *et al.* 1999; Schäfer *et al.* 1999; Deppenmeier 2002). In *Mbb. ruminantium* M1<sup>T</sup>,  $\text{Na}^+$  is utilized by the  $\text{A}_1\text{A}_0$ -ATP synthase for ATP generation, but under conditions of low pH and low  $\text{Na}^+$  a  $\text{H}^+$  gradient could also be used to generate ATP (McMillan *et al.* 2011).

## 1.7 Methanogen Genomes

Advances in next generation DNA sequencing technologies has made genome sequencing a rapid and economical research method (Grada and Weinbrecht 2013). Methanogen genomic information is useful in investigating ruminant  $\text{CH}_4$  mitigation strategies, and one of the strengths of this approach is the ability to compare between genomes to identify which genes are conserved among particular types of methanogens and also which genes define the unique features of each microorganism. Difficulties in the cultivation and purification of rumen methanogens has hindered their widespread genome sequencing. However, the genomes of several cultivated rumen methanogens have been sequenced (Table 1.1). The methanogen genome sequences currently completed range in size from 1.5 Mb to 4.5 Mb, and most tend to be between 1.5 Mb and 3 Mb. The genome of *Methanosarcina barkeri* CM1 is larger than other methanogen genomes. The larger genome size of members of the Methanosarcinales is correlated with their ability to form  $\text{CH}_4$  via hydrogenotrophic, acetoclastic and methylotrophic pathways and thus have more complex metabolic capacities (Maeder *et al.* 2006).

*Methanobrevibacter ruminantium* M1<sup>T</sup> was the first genome sequence of a rumen methanogen (Leahy *et al.* 2010). Like other hydrogenotrophic methanogens, it has a small genome of 2.9 Mb with a %G+C of 33%, and encodes each of the seven steps in the methanogenesis pathway using  $\text{H}_2 + \text{CO}_2$ . In hydrogenotrophic methanogens, the last two steps of the methanogenesis pathway are strongly conserved, which includes the CoM methyltransferase and the methyl CoM reductase genes. Several methanogenesis marker genes, currently without any ascribed function, are also highly conserved (Gao and Gupta 2007; Liu and Whitman 2008). The M1<sup>T</sup> genome sequence explains why it needs acetate, 2-methylbutyrate and CoM to survive, as it lacks several genes involved in the use or synthesis of these compounds. Insights from the analysis of the M1<sup>T</sup> genome have led to the optimization of *Mbb. ruminantium* growth. The presence of two copies of NADP-dependent alcohol dehydrogenase within the genome

suggested possible alcohol utilisation and it was found that methanol and ethanol could stimulate, but not support, growth. Subsequently methanol has been added to media for growth of M1<sup>T</sup>. The M1<sup>T</sup> genome sequence also lead to the discovery of a prophage sequence, designated  $\phi$ mr<sub>u</sub>, and two non-ribosomal peptide synthase (NRPS) genes, which have never been found previously in archaea (Leahy *et al.* 2010). Several more rumen Methanobacteriales genomes have been sequenced, including *Mbb.* sp. AbM4, *Mbb. boviskoreani* JH1<sup>T</sup>, *Mbb. millerae* SM9, *Mbb. olleyae* YLM1, *Mbb. wolinii* SH<sup>T</sup>, *Mbb. millerae* ZA-10<sup>T</sup>, *Methanosphaera* sp. ISO3-F5, *Methanobacterium formicicum* BRM9. Several other rumen methanogen genome sequencing projects are underway as indicated in Table 1.1, including *Mbb. millerae* HW02 and *Mbb.* sp. YE315.

**Table 1.1 Rumen methanogen genomes from Genomes OnLine Database (GOLD).**

Strain Name	GOLD ID	Order	Genome Size (Mb)	GC (%)	ORFs	Completion	Reference
<i>Methanospaera</i> ISO3-F5		Methanobacteriales	2.8	31	2354	In Progress	
<i>Methanobrevibacter ruminantium</i> M1 <sup>T</sup>	Gc01203	Methanobacteriales	2.9	33	2283	Complete	Leahy <i>et al.</i> , 2010
<i>Methanobrevibacter millerae</i> SM9	Gi06992	Methanobacteriales	2.5	32	2321	Complete	Kelly <i>et al.</i> , 2016
<i>Methanobrevibacter</i> YE286		Methanobacteriales				Draft	
<i>Methanobacterium formicicum</i> BRM9	Gi06999	Methanobacteriales	2.4	41	2352	Complete	Kelly <i>et al.</i> , 2014
<i>Methanobacterium</i> YE299		Methanobacteriales				Draft	
<i>Methanobrevibacter olleyae</i> YLM1	Gi07000	Methanobacteriales	2.2	27	1862	Draft	Kelly <i>et al.</i> , 2016
<i>Methanobrevibacter</i> sp. AbM4	Gi17672	Methanobacteriales	2	29	1730	Complete	Leahy <i>et al.</i> , 2013
<i>Methanobrevibacter millerae</i> HW02	Gi0074693	Methanobacteriales				In Progress	
<i>Methanobrevibacter</i> sp. YE315	N/A	Methanobacteriales				In Progress	
<i>Methanobrevibacter wolinii</i> SH <sup>T</sup>	Gi0054485	Methanobacteriales	2	24	1736	Draft	
<i>Methanobrevibacter millerae</i> DSM 16643	Gi0070837	Methanobacteriales	2.7	37	2467	Draft	
<i>Methanobrevibacter boviskoreani</i> JH1 <sup>T</sup>	Gi39333	Methanobacteriales	2.1	29	1774	Draft	Lee <i>et al.</i> , 2013
Thermoplasmatales archaeon BRNA1	Gi0052360, Gc0052360	Methanomassiliicoccales	1.5	58	1577	Complete	
<i>Methanosarcina</i> CM1	Gi06991	Methanosarcinales	4.5	39	3655	Complete	Lambie <i>et al.</i> , 2015

Genomes with N/A (not available) are genome is in early stage of sequencing and no GOLD ID was assigned at current stage. Genomes without a GOLD ID were acquired from (Morgavi *et al.* 2012).

## 1.8 Research questions

Genome sequencing of rumen methanogens provides detailed knowledge of their genetic makeup and provides insights into their metabolism via reconstruction of their metabolic pathways from bioinformatic analyses. However, many different types of methanogens exist within the rumen, and ruminant CH<sub>4</sub> mitigation strategies should target all rumen methanogens and not just a few species. Therefore, in order to assess the genomic diversity of methanogens in the rumen, multiple methanogen genomes selected from all available taxonomic levels (order, family, genera, species and strain), should be sequenced and compared. My PhD project seeks to contribute to this effort by acquiring detailed knowledge of rumen methanogens through genome sequencing and comparative bioinformatic analyses of their genomes.

At the beginning of my PhD project, only the *Mbb. ruminantium* M1<sup>T</sup> genome had been sequenced and was publicly available (Leahy et al 2010) and my interests were in a particular group of recently discovered methanogens called Rumen Cluster C (RCC, now classified as members of the newly formed methanogen order, Methanomassiliicoccales) and in strain level variations in *Methanobrevibacter* genome sequences. A pure culture of RCC had never been isolated from a ruminant, and therefore our knowledge of the rumen RCCs was very limited. The aim of my PhD work was to obtain a wider representation of rumen methanogen genomes by sequencing the DNA of a CH<sub>4</sub>-producing enrichment culture called ISO4-H5, containing a methanogenic archaeon belonging to the RCC group. In addition, sequence *Mbb. gottschalkii* D5 as part of a *Mbb.* ‘pan-genome’ analysis to assess strain level genomic variation in this methanogen species.

My specific research questions are:

- What is the genome composition of the methanogenic archaeon RCC ISO4-H5, and how does this differ from other sequenced methanogens?
- What is the metabolic scheme of RCC ISO4-H5, how does it grow in the rumen, and can these features be used to isolate a pure culture of this organism?
- What is the genome composition of *Methanobrevibacter* sp. D5, how does this differ from other sequenced *Mbb.* strains and how do these differences allow co-existence of multiple *Methanobrevibacter* spp. in the rumen?

## Chapter 2

### Materials and methods

## 2.1 Materials

### 2.1.1 Media

#### 2.2.1.1 Rumen fluid

Rumen fluid was collected from ruminally-fistulated cows fed hay and the rumen fluid was filtered by cheese cloth and stored at -20 °C. Rumen fluid was thawed, boiled, cooled on ice, and the precipitate was removed by centrifugation at  $10,000 \times g$  for 10 min at room temperature in a Sorvall Evolution RC centrifuge (Thermo Fisher Scientific, Waltham, Ma, USA) prior to media preparation.

#### 2.2.1.2 Salt solution A

Salt solution A consists of (g/L) NaCl (6),  $\text{KH}_2\text{PO}_4$  (3),  $(\text{NH}_4)_2\text{SO}_4$  (1.5),  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  (0.79) and  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  (1.2) in distilled water.

#### 2.2.1.3 Salt solution B

Salt solution B consists of (g/L)  $\text{K}_2\text{HPO}_4 \cdot 3\text{H}_2\text{O}$  (7.86) in distilled water.

#### 2.2.1.4 Trace element solution (SL10)

The trace element solution was prepared by addition of 10 mL 25% HCl, 1.5 g  $\text{FeCl}_2 \cdot 4\text{H}_2\text{O}$ , 190 mg  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ , 100 mg  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ , 70 mg  $\text{ZnCl}_2$ , 6 mg  $\text{H}_3\text{BO}_3$ , 36 mg  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ , 24 mg  $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$  and 2 mg  $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$  sequentially into a final volume of 1 L of distilled  $\text{H}_2\text{O}$ , and was sterilized by autoclaving.

#### 2.2.1.5 Selenite/tungstate solution

The selenite/tungstate solution consists of (g/L) NaOH (0.5),  $\text{Na}_2\text{SeO}_3 \cdot 5\text{H}_2\text{O}$  (3 mg) and  $\text{Na}_2\text{WO}_4 \cdot 2\text{H}_2\text{O}$  (4 mg), and was sterilized by autoclaving.

#### 2.2.1.6 BY media

BY medium consists of rumen fluid (30% [v/v]), Salt Solution A (17% [v/v]), Salt Solution B (17% [v/v]), and (g/L)  $\text{NaHCO}_3$  (5g), yeast extract (1), 10 drops of resazurin (0.1% [w/v]), trace element solution SL10 (0.1% [v/v]) (Widdel and Pfennig 1981), selenite/tungstate solution (0.1% [v/v]) (Tschech and Pfennig 1984), and L-cysteine-HCl (0.5). All the components except  $\text{NaHCO}_3$  and L-cysteine-HCl were mixed thoroughly and boiled, bubbled under 100%  $\text{O}_2$ -free  $\text{CO}_2$  for 20 min immediately after boiled while cooled in ice bath,  $\text{NaHCO}_3$

and L-cysteine-HCl were mixed into the cooled media. pH adjustment? The media was dispensed into 100% O<sub>2</sub>-free CO<sub>2</sub>-flushed anaerobic vessels, the vessel was flushed for a minimum of 30 seconds, 15 min and 30 min for Hungate tubes, serum bottles and Schott bottles respectively. For agar-containing media, bacterial agar (1.5% [w/v]) was dispensed into the vessel, flushed with 100% O<sub>2</sub>-free CO<sub>2</sub> prior to media dispensation. The vessels were capped and sealed with headspace of 100% O<sub>2</sub>-free CO<sub>2</sub>, and sterilized by autoclaving for 20 min at 121 °C, and stored in the dark at room temperature. Agar-containing media were held at 55 °C in a water bath post-autoclaving, and additional substrates were added and mixed prior to plating inside an anaerobic hood.

### 2.1.2 Media additives

Unless otherwise mentioned, all media additives were prepared with boiled distilled water, bubbled under 100% O<sub>2</sub>-free N<sub>2</sub> for 10 min.

#### 2.1.2.1 Vitamin 10× concentrate

The Vitamin 10 concentrate stock solution consisted of (mg/L) 4-aminobenzoate (40), d-(+)-biotin (10 mg), nicotinic acid (100) hemicalcium D-(+)-pantothenate (50), pyridoxamine hydrochloride (150), thiamine chloride hydrochloride (100), cyanocobalamin (50), D,L-6,8-thioctic acid (30), riboflavin (30) and folic acid (10). The solution was dispensed anaerobically into sterile, O<sub>2</sub>-free, N<sub>2</sub>-filled serum bottle through 22-μm pore size sterile Millex GP filter (Merck Millipore, Billerica, MA, USA). The bottles were wrapped in aluminium foil to protect light sensitive vitamins and stored at 4 °C for short term or -20 °C for long term. The working stock was diluted 10 times with growth medium.

#### 2.1.2.2 Sodium formate (3M)

Sodium formate (204.03 g/L) was dissolved in 100% O<sub>2</sub>-free water, bubbled with 100% O<sub>2</sub> free N<sub>2</sub> for 20 min, sealed with 100% O<sub>2</sub>-free N<sub>2</sub> headspace, and sterilized by autoclaving.

#### 2.1.2.3 Sodium acetate (1M)

Sodium acetate.3H<sub>2</sub>O (136.08 g/L) was dissolved in O<sub>2</sub> free N<sub>2</sub> water, bubbled with 100% O<sub>2</sub> free N<sub>2</sub> for 20 min, sealed with 100% O<sub>2</sub>-free N<sub>2</sub> headspace, and sterilized by autoclaving.

#### 2.1.2.4 Methanol (1 M)/ethanol (1 M)/butanol (1 M)

Methanol (41 mL/L, [v/v]) or analytical grade ethanol (58.4 mL/L, [v/v]) or butanol (91.5 mL/L, [v/v]) was mixed in O<sub>2</sub> free N<sub>2</sub> water, bubbled with 100% O<sub>2</sub> free N<sub>2</sub> for 20 min,



dispensed anaerobically into sterile, O<sub>2</sub> free N<sub>2</sub> filled serum bottle through 22 µm pore size sterile Millex GP filter (Merck Millipore).

#### *2.1.2.5 Glucose (1 M)*

Glucose (180.16 g/L) was dissolved in O<sub>2</sub>-free water, bubbled with O<sub>2</sub>-free N<sub>2</sub> for 20 min, dispensed anaerobically into sterile, O<sub>2</sub>-free N<sub>2</sub>-filled serum bottles through 22 µm pore size sterile filter (Merck Millipore).

#### *2.1.2.6 CoM (10 mM)*

CoM (1.41 g/L) was dissolved in O<sub>2</sub>-free water, bubbled with O<sub>2</sub>-free N<sub>2</sub> for 20 min, dispensed anaerobically into sterile, O<sub>2</sub>-free N<sub>2</sub>-filled serum bottles through 22 µm pore size sterile filter (Merck Millipore), and stored in the dark at 4 °C.

#### *2.1.2.7 Pectin (10% [w/v])*

Apple pectin (100 g/L) was dissolved in O<sub>2</sub>-free water, bubbled with O<sub>2</sub>-free N<sub>2</sub> for 20 min, sealed with 100% N<sub>2</sub> headspace, sterilized by autoclaving and stored at 4 °C. The solution was allowed to stand for a period (≥12 h) to dissolve.

#### *2.1.2.8 Cellulose (5%, [w/v])*

Cellulose paper was cut to small pieces with scissors and ground to a powder in distilled water in a ball mill at 4 °C, bubbled with O<sub>2</sub>-free N<sub>2</sub> for 20 min, sealed in serum bottles with 100% N<sub>2</sub> headspace, and sterilized by autoclaving.

#### *2.1.2.8 Antibiotics*

Ampicillin (50 mg/mL, Sigma-Aldrich, St. Louis, MO, USA), gentamycin (20 mg/mL, Sigma-Aldrich, St. Louis, MO, USA), kanamycin (50 mg/mL, Sigma-Aldrich, St. Louis, MO, USA), spectinomycin (100 mg/mL, Sigma-Aldrich, St. Louis, MO, USA) and streptomycin (100 mg/mL, Sigma-Aldrich, St. Louis, MO, USA) were dissolved in O<sub>2</sub>-free water. Chloramphenicol (50 mg/mL, Sigma-Aldrich, St. Louis, MO, USA), erythromycin (20 mg/mL, Sigma-Aldrich, St. Louis, MO, USA), tetracycline (10 mg/mL, Sigma-Aldrich, St. Louis, MO, USA) was dissolved in N<sub>2</sub>-bubbled ethanol. Rifampicin (12 mg/mL, Sigma-Aldrich, St. Louis, MO, USA) was dissolved in N<sub>2</sub>-bubbled methanol. Antibiotics were sterilized by filtration through 22 µm pore size sterile filter (Merck Millipore), sealed with 100% N<sub>2</sub> headspace and stored at -20 °C, tetracycline and rifampicin were stored in the dark.

#### 2.1.2.9 Dimethyl sulfide (1 M)/Methyl 3-mercaptopropionate (M3SP, 1 M)/methyl-3-methylthiopropionate (M3MSP, 1 M)

Dimethyl sulfide (73.9 mL/l, [v/v], Sigma-Aldrich)/ M3SP (110.8 mL/l, [v/v], Sigma-Aldrich)/ M3MSP (124.61 mL/l, [v/v], Sigma-Aldrich) were injected anaerobically into autoclaved BY media with 100% CO<sub>2</sub> headspace.

### 2.1.3 Methanogen and bacteria strains

#### 2.1.3.1 Methanogenic archaeon ISO4-H5 enrichment culture and *Succinivibrio dextrinosolvens* H5

An enrichment culture of a representative of the Methanomassiliicoccales was previously obtained from the rumen contents of a 9 year old Romney wether sheep in NZ grazing a ryegrass-clover pasture diet (Jeyanathan 2010). This enrichment culture contained a methanogenic archaeon, designated ISO4-H5 and a Gram-negative bacterium, subsequently identified as being most closely related to *Succinivibrio dextrinosolvens*, and designated as strain H5 (Jeyanathan 2010).

#### 2.1.3.2 *Methanobrevibacter* sp. D5 and enrichment culture H6

A methane-forming enrichment culture H6 was originally obtained from the rumen of sheep J472 on pasture diet (G. Henderson, personal communication). A member of the *Mbb. gottschalkii* clade was isolated from the H6 culture and designated as *Methanobrevibacter* sp. D5 (Kim 2012).

### 2.1.4 General materials

#### 2.1.4.1 PCR reagents

Type I water used for PCR was purified by Milli-Q Water purification system (Merck Millipore, Darmstadt, Germany). All PCR amplifications were performed in Mastercycler<sup>TM</sup> Pro PCR system (Eppendorf, Hamburg, Germany) using Platinum® PCR supermix or Platinum® PCR supermix high fidelity (Thermo Fisher Scientific, Waltham, Ma, USA) for large products.

#### 2.1.4.2 Primers

Primers were purchased from Integrated DNA Technologies (Integrated DNA technologies Coralville, IA, USA). The stock solutions are made up to 200 µM with Type I water and stored at -20 °C. The stock solution was diluted to working concentration of 20 µM with Type I water when required.

#### 2.1.4.3 50×TAE buffer

50×TAE buffer consists of 2 M Tris, 1 M acetic acid and 50 mM pH 8.0 ethylenediaminetetraacetic acid (EDTA). 1× TAE buffer is prepared by dilution with distilled water.

#### 2.1.4.4 10×TBS buffer

10×TBS buffer consists of Tris (54 g/L), boric acid (27.5 g/L) and 10 mM pH 8.0 EDTA. 0.5× TBE buffer was prepared by dilution with distilled water.

#### 2.1.4.5 100× TE buffer

100× TE buffer consists of 1 M Tris-HCl, 100 mM EDTA, pH adjusted with HCl/NaOH to suit experimental requirement. The pH range used in this thesis includes pH 7.6 and pH 8.0. 1× TE buffer was prepared by dilution with Type I water.

### 2.1.5 RNA purification materials

Zirconia/Silica beads (dnature Ltd, Gisborne, New Zealand) of 0.1 mm and 0.5 mm diameter (1:1) and glassware were baked at 180 °C for two hours to sterilize and deactivate RNase. All surface and equipments were decontaminated by RnaseZap decontamination solution (Thermo Fisher Scientific) prior experiment. RNase free filtered tips were used. Unless otherwise stated, water used is Nuclease-free water, Ambion® (Thermo Fisher Scientific). Diethylpyrocarbonate (DEPC) treated water is prepared by addition of 0.1% [v/v] DEPC to Type I water and incubated for two hours at 37 °C and autoclaved. RNA extraction buffer A consisted of 200 mM NaCl, 20 mM EDTA (pH 8.0) in Type I water and DEPC treated. 20% [w/v] SDS was dissolved in Type I water and DEPC treated. RNase-free ammonium acetate (5M), Ambion® and phenol:chloroform:isoamyl alcohol (125:24:1, pH 4.5) was supplied by Thermo Fisher Scientific. Purification of RNA was carried out with MEGAclean Transcription Clean-Up Kit (Thermo Fisher Scientific), DNA removal was carried out with Baseline-ZERO DNase (Illumina Netherlands BV, Eindhoven, Netherlands) and TURBO DNA-free Kit (Thermo Fisher Scientific).

## 2.2 Methods

### 2.2.1 Anaerobic cultivation

All methanogens used in this study were cultivated anaerobically in BY media. The cultivation and media preparation techniques were introduced by Robert Hungate (Hungate 1966) with

adaptations by Balch (Balch and Wolfe 1976). The medium was dispersed into vessels flushed by O<sub>2</sub>-free CO<sub>2</sub>. Vessels used in this study include 16 × 125 mm Hungate tubes (BellCo Glass, Vineland, NJ, USA), high pressure Schott bottles with open top screw caps and butyl rubber stoppers and 125 mL serum bottles (BellCo) with butyl rubber stopper. The media was sterilised using a TOMY SX-700E autoclave (Digital Biology, Tokyo, Japan). Filter-sterilized stock solutions were introduced into media anaerobically and aseptically post-autoclaving.

#### 2.2.1.1 *Succinivibrio Spent Pectin Growth Medium Supernatant (SSPGMS)*

Pectin solution (1%, [w/v]) was added to the pure culture of *Succinivibrio dextrinosolvens* H5 in BY medium, cultivated for two days at 39 °C and stored at -20 °C. The pectin-incubated culture was thawed and dispensed anaerobically into ISO4-H5 cultures through 22 µm pore size sterile filters (Merck Millipore).

#### 2.2.1.2 *Growth measurement*

Methanogen growth was monitored by methane production. The head space gas (0.3 mL) was taken by syringe and measured by gas chromatography using an Aerograph 660 (Varian Associates, Palo Alto, Ca, USA) fitted with a Porapak Q 80/100 mesh column (Waters Corporation, Milford, MA, USA) and a thermal conductivity detector operated at 100 °C. The column used N<sub>2</sub> as the carrier gas at 12 cm<sup>3</sup>/min at room temperature. Standard curves were constructed for H<sub>2</sub> and CH<sub>4</sub> at standard atmospheric pressure. Bacterial growth was monitored by optical density at 600 nm within Hungate tubes via a spectrophotometer (Spectronic 200, Thermo Fisher Scientific).

#### 2.2.1.3 *Standard sub-culture conditions*

The required substrates and supplements were added to the medium, then inoculated with 10% of culture and pressurized with high-pressure H<sub>2</sub>:CO<sub>2</sub> (4:1, 180 kPa). The maximum period of culture viability is approximately two weeks for methanogenic archaeon ISO4-H5 and *S. dextrinosolvens* H5, and one week for *Methanobrevibacter* sp. D5. Culture viability was achieved by sub-culturing to fresh medium. An additional tube of culture with glucose (10 mM) was used to identify contamination in the ISO4-H5 pure culture and *Methanobrevibacter* sp. D5 culture with microscopy of culture wet mounts.

**Table 2.1** The final concentrations or volumes of additives to a 10 mL culture of BY medium in Hungate tubes.

	Methanogenic archaeon ISO4-H5 enrichment culture	Methanogenic archaeon ISO4-H5 pure culture	<i>Succinivibrio dextrinosolvens</i> H5	<i>Methanobrevibacter</i> sp. D5
Sodium formate	60 mM	60 mM	-	60 mM
Sodium acetate	20 mM	20 mM	-	20 mM
Methanol	20 mM	20 mM	-	20 mM
Vitamin mix (working stock)	0.1 mL	0.1 mL	-	0.1 mL
CoM	10 $\mu$ M	10 $\mu$ M	-	10 $\mu$ M
Glucose	-	-	10 mM	-
SSPGMS*	-	1 mL	-	-

\* improved growth, not mandatory for culture survival

## 2.2.2 Microscopy

### 2.2.2.1 Light microscopy

A drop of culture was placed onto the microscope slide and covered with a cover slip. Excess liquid was removed by pressing the cover slip with a paper towel. The *Methanobrevibacter* sp. D5 was observed by fluorescent microscopy (420 nm) in combination with phase contrast microscopy (DM2500 microscope, Leica Microsystems, Wetzlar, Germany). Methanogenic archaeon ISO4-H5 was observed under phase contrast microscopy only, as ISO4-H5 has not been observed to fluorescence.

### 2.2.2.2 Transmission electron microscopy

Electron microscopy (Philips CM10 Transmission Electron Microscope with SIS Morada high-resolution digital imaging, Netherlands) was carried out at the Manawatu Microscopy and Imaging Centre (MMIC), Massey University, Palmerston North, New Zealand.

*Negative staining transmission electron microscopy (TEM).* The procedure is identical to previously described (Jeyanathan 2010) with minor adaptation. Cells were collected by centrifugation ( $5,000 \times g$ , 5 min) and resuspended with anaerobic water. One drop of cell suspension was placed onto carbon-coated formvar film attached to copper grids and left for 2-6 min. Excess sample was drained from grids with filter paper. The grids were placed on a drop of 2% uranyl acetate (in water) resting on parafilm and left for 6-10 min. Excess liquid was drained well and allowed to dry. The whole cells were then observed by electron microscopy at 60 Kv.

*2.2.2.3 Thin section transmission electron microscopy.* Cell pellet was collected by centrifugation ( $5,000 \times g$ , 5 min), the cell pellet was fixed by 3% (v/v) glutaraldehyde in 0.1 M  $\text{Na}_2\text{HPO}_4$ ,  $\text{KH}_2\text{PO}_4$  buffer (pH 7.2) and letting it stand for 2 h at room temperature. The cell

suspension was pelleted by centrifugation ( $4,500 \times g$ , 5 min). The cells were rinsed in 0.1 M  $\text{Na}_2\text{HPO}_4$ ,  $\text{KH}_2\text{PO}_4$  buffer three times for 5, 10 and 15 min, and fixed in 1% (w/v) osmium tetroxide ( $\text{OsO}_4$ ) in the same buffer for 1 h at room temperature. The cells were rinsed again in 0.1 M  $\text{Na}_2\text{HPO}_4$ ,  $\text{KH}_2\text{PO}_4$  buffer three times for 5, 10 and 15 min, followed by dehydration in increasing concentrations of acetone consisting of 25% (v/v), 50% (v/v), 75% (v/v), 95% (v/v) and three changes at 100% (v/v), each step was for 15 min. The dehydrated cells were mixed with 50/50 acetone/resin (Procure 812: ProSciTech, Qld, Australia) (v/v) on a stirrer overnight to allow the acetone to evaporate slowly. Cells were embedded in fresh resin (100%) for 8 h, overnight and then 8 h on a stirrer, and then polymerized at 60 °C for 48 h in an oven. The embedded cell blocks were sectioned with a diamond knife on an Ultra- microtome (Leica Microsystems, Wetzlar, Germany), and collected onto copper grids, stained with saturated uranyl acetate in 50% (v/v) ethanol for 4 min and lead citrate for 4 min. The stained sections were observed by electron microscopy at 60 kV.

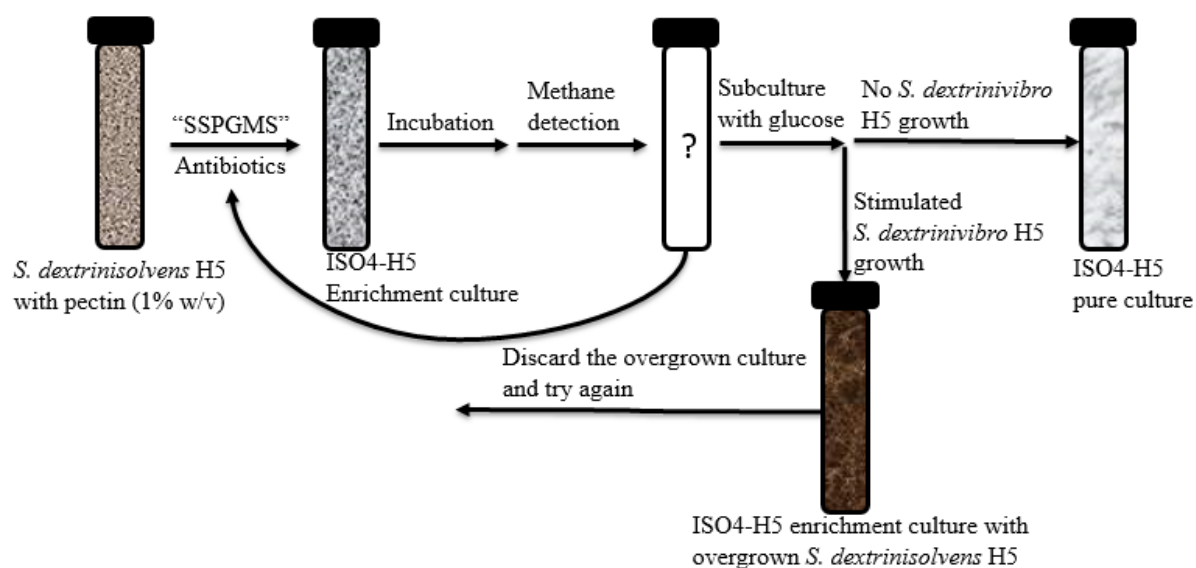
### 2.2.3 Purification of methanogens

#### 2.2.3.1 Minimum inhibitory concentration of antibiotics against *Succinivibrio dextrinsolvens* H5 within ISO4-H5 enrichment culture

Ampicillin (50 µg/mL), chloramphenicol (50 µg/mL), erythromycin (20 µg/mL), gentamycin (20 µg/mL), kanamycin (50 µg/mL), rifampicin (1.2 µg/mL), spectinomycin (100 µg/mL), streptomycin (100 µg/mL) or tetracycline (10 µg/mL) were supplemented into triplicates of ISO4-H5 enrichment culture in the presence of glucose (10 mM), and the growth of *S. dextrinsolvens* H5 was compared to triplicates of negative control omitting antibiotics after overnight incubation (16 h). The  $\text{CH}_4$  production was measured after seven days of growth with triplicates of ISO4-H5 enrichment culture under standard cultivation conditions as a positive control. The concentration of antibiotics was increased to identify the minimum inhibitory concentration that did not compromise methane production.

#### 2.2.3.2 Purification of Methanogenic archaeon ISO4-H5

Methanogenic archaeon ISO4-H5 cultures supplemented with SSPGMS were treated with ampicillin (200 µg/mL), sub-cultured into BY medium supplemented with glucose (10 mM) to test for surviving *S. dextrinsolvens* H5, or media supplemented with SSPGMS and ampicillin to remove *S. dextrinsolvens* H5 (Figure 2.1). This procedure was repeated until no bacterial contamination was detected.



**Figure 2.1 Isolation procedure for ISO4-H5.**SSPGMS: *Succinivibrio* Spent Pectin Growth Medium Supernatant. Antibiotics: 200 µg/ml ampicillin. Glucose: 10 mM. Length of incubation: one week.

#### 2.2.3.3 Purification of *Methanobrevibacter* sp. D5

*Methanobrevibacter* sp. D5 cultures with bacterial contamination were subjected to a mixture of ampicillin (10 µg/mL), streptomycin (10 µg/mL) and vancomycin (86.7 µg/mL), ten heat-treated at 55°C for 30 min followed by streaking on anaerobic BY agar plates. The agar plates were placed in a sealed cannister and pressurised with 1 atm over pressure of H<sub>2</sub>:CO<sub>2</sub> (4:1). Plates were incubated at 39 °C for one week, and a pure *Methanobrevibacter* sp. D5 culture was isolated from a single colony and the subsequent culture purity was verified by 16S rRNA gene sequencing.

#### 2.2.4 DNA extraction

##### 2.2.4.1 Genomic DNA extraction

The protocol was adapted from the Genomic-Tip 500/G kit (Qiagen, Silicon Valley, Redwood city, CA, USA). A mortar and pestle were sterilized and stored at -80 °C. A 1 L culture of *Methanobrevibacter* sp. D5 was grown in Schott bottles or serum bottles. Culture purity was validated by microscopy. Cells were collected by centrifugation (20,200 × g, 20 min) in sterile centrifuge bottles and Oakridge tubes. The cell pellet was froze at -80 °C for a minimum of two hours.

The cell pellet was transferred to the pre-chilled mortar and crushed with the pestle whilst kept frozen with liquid N<sub>2</sub>. The pellet was ground in a circular motion as it slowly thawed into a slushy consistency, and additional liquid nitrogen was added to keep the pellet from thawing completely. RNase A (2 µg/mL, Qiagen) was added to 11 mL of Buffer B1 solution (Qiagen).

The ground cells were resuspended with 5 mL of Buffer B1, using a wide-bore pipette tip. The lysate was transferred to a 50 mL centrifuge tube (Falcon™, Fisher Scientific International Inc. Hampton, USA), and 6 mL of additional buffer B1 was added. Freshly prepared lysozyme (2.5 mg/mL [w/v], Sigma-Aldrich, St. Louis, MO, USA) and proteinase K (8 mg/mL [w/v], Qiagen) was added to the lysate mixture. The remainder of the purification procedures followed the Qiagen Genomic DNA Handbook for bacterial genomic DNA extraction.

The DNA extraction procedure of methanogenic archaeon ISO4-H5 was similar to the procedure used for *Methanobrevibacter* sp. D5, except due to low cell density, 2 L of ISO4-H5 culture was grown. ISO4-H5 cultivation was possible in 125 mL serum bottles but not in 250 mL, 500 mL or 1 L Schott bottles. ISO4-H5 did not require grinding, the addition of SDS (0.1%, [w/v]) to the lysate buffer along with proteinase K and lysozyme was sufficient to release genomic DNA.

#### 2.2.4.2 DNA extraction for 16S rRNA polymerase chain reaction

DNA was extracted from pellets of cells using the ZR Fungal/Bacterial DNA MiniPrep™ Kit (Zymo Research Corporation, Irvine, CA, USA) with minor modification to the instruction. The bead beating time was shortened to 90 seconds.

#### 2.2.5 Polymerase chain reaction (PCR)

Methanogen genome gap regions and partial fragments of 16S rRNA gene from both archaea and bacteria were amplified in this study. Each PCR (50 µL) reaction contained 45 µL of platinum supermix, 0.4 µM of each primer, 5 ng of template DNA and Type I water. Each set of PCR contained a positive control with sequence validated template for 16S rRNA gene amplification, a negative control replacing template with Type I water was used to identify false positive reactions and primer dimer formation. The primers used in this study are listed in Tables 2.2, A.2.1 and A.2.2.

**Table 2.2 Primers used for PCR amplification of 16S rRNA gene**

Primer pair	Target gene (fragment size)	Sequence (5' – 3')	Thermal cycling conditions	Reference
27f 1492r	Bacterial 16S rRNA (≈1465 bp)	GAG TTT GAT CMT GGC TCA G GGY TAC CTT GTT ACG ACT T	94°C 10 min 94°C 30 s 56°C 30 s 72°C 90 s 72°C 10 min } 30 cycles	Weisburg <i>et al.</i> 1991
915af 1386r	Archaeal 16S rRNA (≈471 bp)	AGG AAT TGG CGG GGg AGC AC GCG GTG TGT GCA AGG AGC	94°C 10 min 94°C 30 s 56°C 30 s 72°C 30 s 72°C 10 min } 30 cycles	Watanabe <i>et al.</i> 2004, Skillman <i>et al.</i> 2004



**Table 2.3 Primers used for PCR amplification of ISO4-H5 putative prophage region**

Primer pair	Target gene and fragment size	Sequence (5' – 3')	Thermal cycling conditions	Reference
H5PHX.p1ca	Putative prophage full length (32.8 kb) or ISO4-H5 genome (≥500bp)	GTC CTC CAA GGG TTA TTC	94°C 10 min	This study
H5PHX.q1ca		GT	94°C 30 s	
		TCG TTT CCG GTT AAA AGA	56°C 30 s	
		CG	68°C 10 min 68°C 10 min	
H5PH1.p1ca	Putative prophage internal region (470 bp)	CTG CTA TAC ATA TTG TGT	94°C 10 min	This study
H5PH1.q1ca		CG	94°C 30 s	
		TCA TTT CTT CTT CGT AGA	56°C 30 s	
		GG	72°C 60 s 72°C 10 min	
H5PH2.p1ca	Putative prophage internal region (872 bp)	CTG CAG ACC TGT TCT GAC	94°C 10 min	This study
H5PH2.q1ca		TA	94°C 30 s	
		ATC AAG GTG GAG AAA TTT	56°C 30 s	
		GG	72°C 60 s 72°C 10 min	
H5PH3.p1ca	Putative prophage internal region (924 bp)	TCG GCC TAT TAG CCT CTC	94°C 10 min	This study
H5PH3.q1ca		TTT CAG	94°C 30 s	
		CAA AAG GTC CGG CAA AAT	56°C 30 s	
		CC	72°C 60 s 72°C 10 min	
H5PH4.p1ca	Putative prophage internal region (904 bp)	TCT GGT CAT CAC CTC GAT	94°C 10 min	This study
H5PH4.q1ca		TG	94°C 30 s	
		TTC CGC TTC TGC CTA CGA	56°C 30 s	
		GA	72°C 60 s 72°C 10 min	
H5PH5.p1ca	Putative prophage internal region (850 bp)	GCT GGA TCT CGC CGA CTG	94°C 10 min	This study
H5PH5.q1ca		AT	94°C 30 s	
		TCG AGA TCG TTT CTT CCA	56°C 30 s	
		CC	72°C 60 s 72°C 10 min	

## 2.2.6 Agarose gel electrophoresis

Amplified genome gap regions or 16S rRNA gene products were analysed by agarose gel electrophoresis. Agarose gels (1% [w/v]) were made up in 1× TAE buffer. Amplified PCR products were loaded with SYBR Safe DNA Gel Stain (20%, [v/v], Thermo Fisher Scientific) and run in 1× TAE buffer at 100 V for 30 min. 1Kb<sup>+</sup> DNA ladder (Thermo Fisher Scientific) was used as molecular weight marker in each gel. The gel was visualised under ultraviolet (UV) irradiation using a Gel Logic 200 imaging system (Eastman Kodak, New York, NY, USA).

## 2.2.7 PCR product purification

All PCR products were purified by ExoStar 1-Step (29%, [v/v], GE Healthcare, Little Chalfont, United Kingdom) by incubation at 37 °C for 15 min and 80 °C for 15 min.

## 2.2.8 DNA quantification

The DNA was quantified by using NanoDrop ND-1000 UV-Vis Spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA) according to the manufacturer's instructions.

## 2.2.9 Sequencing of PCR products

PCR products were sequenced by Massey Genome Service (Massey University, Palmerston North, New Zealand). Chromatogram of sequencing result were visualised and edited using Geneious 6.1.5 (Kearse *et al.* 2012).

## 2.2.10 Genome sequencing

### 2.2.10.1 *Methanogenic archaeon ISO4-H5 genome*

Genomic DNA from the methanogenic archaeon ISO4-H5 enrichment culture was extracted as described in Section 2.2.1. The paired-end library with a minimum insert size of 3 kb was prepared and whole-genome sequencing was carried out on a 454 GS-FLX Titanium System pyrosequencing platform (Macrogen, Seoul, South Korea).

### 2.2.10.2 *Methanobrevibacter sp. D5 genome*

Genomic DNA from a methanogen enrichment culture H6 containing *Methanobrevibacter sp. D5* was determined by next-generation sequencing on a 454 GS-FLX Titanium System pyrosequencing platform (Macrogen, Seoul, South Korea). *Methanobrevibacter sp. D5* was subsequently purified as described in Section 2.2.1, and genomic DNA was extracted as described in Section 2.8.1. Genomic DNA from a pure *Methanobrevibacter sp. D5* culture was sequenced again using an Illumina MiSeq pair-end reads with a 250 bp insert size (Beijing Genomics Institute, Yantian District, Shenzhen, China).

## 2.2.11 Quality assessment

The quality of the raw DNA reads were validated by FastQC (Andrews 2010) and poor quality reads were trimmed with FLEXBAR (Dodt *et al.* 2012) using a Phred-scale cut-off of 33 and a minimum read size of 100 bp prior to genome assembly.

```
“Flexbar -r D5_1.fastq -p D5_2.fastq -n 4 -a adap.all.fa -q 33 -f i1.8 -m 100”
```

### 2.2.11.1 *Methanogenic archaeon ISO4-H5 genome*

Trimmed reads were assembled by the Genome Sequence de novo Assembler (Newbler) version 2.7 (Margulies *et al.* 2005) developed by Roche. The infoseq tool of the EMBOSS software suite was used to assess %G+C of each contigs (Rice *et al.* 2000). The *mcrA* gene of *Mbb. ruminantium* M1<sup>T</sup> was used as a reference to identify methanogenesis marker genes via BLASTN. The contigs and scaffolds were assigned to ISO4-H5 or *S. dextrinosolvens* H5 based on difference in G+C% and marker genes for methanogenesis. The command used in this study was:

```
“infoseq contigs.fa > assembly.txt”
```

#### 2.2.11.2 *Methanobrevibacter* sp. D5 genome

The trimmed 454 reads of methanogen enrichment culture H6 was assembled by the Genome Sequence de novo Assembler (Newbler) version 2.7 (Margulies *et al.* 2005). The contigs and scaffolds were assigned to *Methanobrevibacter* sp. D5 based on difference in %G+C, identity of ribosomal proteins, presence of methanogenesis marker genes and the tetra scores calculated by TETRA software in relation to scaffold one (Teeling *et al.* 2004), which was the largest *Methanobrevibacter* sp. D5 scaffold. The reads of *Methanobrevibacter* sp. D5 pure culture was assembled by SPAdes (Bankevich *et al.* 2012). The command used in this study was:

```
“Python bin/spades.py --pe1-1 ../D52/D5_1.fastq --pe1-2 ../D52/D5_2.fastq -o spades_output”
```

#### 2.2.12 Genome Annotation

The genome annotation was managed by a combination of GAMOLA (Altermann and Klaenhammer 2003) and Artemis software suite (Rutherford *et al.* 2000). The open reading frames (ORFs) and gene identities were predicted by Glimmer (Delcher *et al.* 1999) and BLASTX (Gish and States 1993). The start and stop of predicted ORFs were verified by manual inspection and redefined if necessary. The protein function of ORFs was assigned manually based on results from databases including: BLASTP (Altschul *et al.* 1990) to both non-redundant protein database provided by the National Centre for Biotechnology Information (NCBI) (Sayers *et al.* 2010), clusters of orthologous groups (COG) database (Tatusov *et al.* 2001). The protein motifs were identified by HMMER (Eddy 1998) search against the Pfam (Finn *et al.* 2016) and TIGRFAM (Haft *et al.* 2003) database.

#### 2.2.13 Primer design

Primers were designed in the Staden Package using the Gap4 software (Staden *et al.* 2000). The primers were designed with a minimum length of 20 bp to ensure uniqueness. The viability of primer sequence was validated in Vector NTI 11.5.2 (Thermo Fisher Scientific), using a cutoff of -4.7 kcal/mol and a maximum of 5 bp stem length to prevent primer dimer formation.

#### 2.2.14 Genome gap closure

The genomes were closed by PCR-based gap closure. The gap regions within scaffolds were amplified with primers designed using neighbouring contigs as templates. Small contigs below 1,000 bp were not used as template due to low reliability. The gap regions between scaffolds were amplified with combination PCRs to test all possible connections between scaffolds. The gap closure progress was managed by the Gap4 software of the Staden Package (Staden *et al.*

2000). The potential frameshifts were identified, and assessed by sequencing, and were removed when found.

#### 2.2.15 Pulse Field Gel Electrophoresis (PFGE)

The PFGE was performed using a CHEF-DR II Pulsed Field Electrophoresis System (Bio-rad, Hercules, CA, USA) to verify genome assembly and scaffold orientation. The procedure was identical between the *Methanobrevibacter* sp. D5 and that previously described for *Mbb. ruminantium* M1<sup>T</sup> PFGE (Leahy *et al.* 2010). Cells were harvested from 50 mL of culture after seven days, by centrifugation ( $10,000 \times g$ , 10 min), frozen with liquid N<sub>2</sub>, gently ground by mortar and pestle to damage the cell wall, and thawed. The ground material was suspended in 2 mL of 1 M NaCl, 10 mM Tris (pH 7.6), and 300  $\mu$ L aliquots were mixed with equal volumes of 2% (w/v) Seaken® Gold low melt agarose (Lonza Group AG, Basel, Switzerland) to embed cells. The cells were lysed in lysis buffer (50 mM Tris-HCl, 50 mM EDTA, 1% [w/v] sarkosyl, 0.1 mg/mL Proteinase K (Qiagen)) at 50 °C for an hour, then washed twice with autoclaved Type I water and three times with TE buffer (pH 8.0), before storing the plug in 10 mM Tris-HCl, 100 mM EDTA (pH 8.0) at 4 °C. The restriction enzymes *Apa*I and *Mlu*I (New England Biolabs, Beverly, MA, USA) were used to digest the embedded *Methanobrevibacter* sp. D5 genomic DNA. The embedded DNAs were loaded into 1% (w/v) Seaken Gold agarose, run at 200 V, 20 h, 14°C in 0.5×TBE buffer, The gel was stained with ethidium bromide and visualised using a Gel Doc 1000 system (Kodak Gel Logic 200 Imaging System, Eastman Kodak, Rochester, NY, USA).

#### 2.2.16 Inference of phylogenetic divergence

##### 2.2.16.1 Functional genome distribution (FGD) tree

The similarity between genomes can be inferred from an organism's ORFeome using the UPGMA method (Sneath and Sokal 1962) based on the FGD dissimilarity matrix (Altermann 2012).

##### 2.2.16.2 Phylogenic trees based on 16S rRNA gene

The 16S rRNA gene DNA sequence were used to infer the phylogenetic relationship of the methanogens analysed in this thesis. Evolutionary analysis were conducted in MEGA6 (Tamura *et al.* 2013). The initial tree for the heuristic search was obtained by applying the Neighbour-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The substitution model was based on the Kimura 2-parameter distance (Kimura 1980). All positions containing gaps or missing data were

eliminated. The bootstrap consensus tree inferred from 1000 replicates (Felsenstein 1985) was used to infer the relationship of the taxa analysed.

#### *2.2.16.3 Phylogenetic trees based on mcrA/mrtA amino acid sequence*

The highly conserved methanogenesis marker gene *mcrA/mrtA* amino acid sequence were used to infer the taxonomic relationship of methanogens analysed. Evolutionary analysis were conducted in MEGA6 (Tamura *et al.* 2013). The initial tree for the heuristic search was obtained by applying the Neighbour-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The substitution model was based on the Jones-Taylor-Thornton (JTT) distance matrix (Jones *et al.* 1992) at uniform rates. All positions containing gaps or missing data were eliminated. The bootstrap consensus tree inferred from 1000 replicates (Felsenstein 1985) was used to infer the relationship of the taxa analysed.

#### *2.2.16.4 Pangenome tree based on gene families*

Pan genome tree of was inferred based on the absence and presence of gene families by CMG-Biotools (Vesth *et al.* 2013). The relative Manhattan distance were calculated by hierarchical clustering. The bootstrap consensus tree inferred from 1000 replicates (Felsenstein 1985) was used to infer the evolutionary history of the taxa analysed.

### *2.2.17 Prediction of RNAs*

#### *2.2.17.1 Prediction of tRNAs*

The program tRNAscan-SE 1.3.1 was used to detect tRNA genes in genomic DNA, which utilised the Cove probabilistic RNA prediction package (Eddy and Durbin 1994) with improved algorithm implemented for searching for eukaryotic tRNA promoters (Lowe and Eddy 1997). Cove probabilistic RNA prediction maximized the sensitivity, and combined with the ‘archaeal’ option allowed the identification of otherwise undetected archaeal tRNAs that contains introns at non-canonical position. The command used in this study was:

“tRNAscan-SE -o output.txt -m outputstats.txt -C -A input.fasta”.

#### *2.2.17.2 Prediction of non-coding RNAs*

The program Infernal (“INFERence of RNA ALignment”) 1.1 was used to predict non-coding RNAs within the genomic DNA. Infernal utilises covariance models to identify RNA homologs with secondary structure conservation to non-coding RNAs in reference databases (Nawrocki

and Eddy 2013). The Rfam database was used as reference (Nawrocki *et al.* 2015). A cut off of 1e-05 was applied to remove false positives. The command used in this study was:

“cmscan Rfam.cm input.fa > output.txt”

#### 2.2.18 Prediction of signal peptides and secretome

Signal peptides were predicted by SignalP 4.1 (Petersen *et al.* 2011) ([www.cbs.dtu.dk/services/SignalP](http://www.cbs.dtu.dk/services/SignalP)). The SignalP-HMM algorithm trained on eukaryotes, Gram positive and Gram negative bacteria were applied to identify signal peptides in Archaeal genomes. The transmembrane helices were predicted by TMHMM (Krogh *et al.* 2001) using default parameters. The lipobox was predicted by LIPO (Berven *et al.* 2006) using default parameters. The LPxTG-like motif was predicted using CW-PRED (Fimereli *et al.* 2012) using default parameters.

#### 2.2.19 Prediction of adhesin-like proteins

In order to identify adhesin-like proteins in the genome, an integrated computational approach was used. Adhesin prediction was carried out with SPAAN (Sachdeva *et al.* 2005), a P<sub>ad</sub>-value was generated as predicted likelihood of genes being adhesin. Additional subcellular localization algorithms LOCTree3, PSORTb and SubLoc were included to minimize false positive (Chen *et al.* 2006; Yu *et al.* 2010; Goldberg *et al.* 2014). Proteins with a SPAAN P<sub>ad</sub>-value of 0.70 or above and two of the subcellular localization algorithm predicted as extracellular, secreted, plasma membrane or cellwall were considered as adhesin-like proteins.

#### 2.2.20 Prediction of horizontal transfer regions and genes

In order to investigate horizontal gene transfer, Alien\_Hunter (Vernikos and Parkhill 2006), an annotation-independent analysis software that utilises interpolated variable order motifs (IVOM) was used. The disadvantage of this parametric method is the high number of false positives (Azad and Lawrence 2007). The regions with abnormal %G+C and codon composition are more likely to be horizontally transferred. The codon composition is indicated by codon adaptation index (CAI) (Sharp and Li 1987) calculated by CAIcal (Puigbo *et al.* 2008) (<http://genomes.urv.cat/CAIcal/>).

**Table 2.4 Bioinformatics software used in this study**

Software	Description/Application	Source	Reference
Genome sequence de novo assembler 2.7	Genome assembler for 454 sequencing platform	<a href="http://www.454.com/products/analysis-software/">http://www.454.com/products/analysis-software/</a>	Margulies <i>et al.</i> 2005
Velvet de novo assembler	Genome assembler for Illumina sequencing platform	<a href="https://www.ebi.ac.uk/~zerbino/velvet/">https://www.ebi.ac.uk/~zerbino/velvet/</a>	Zerbino 2010
SPAdes	Genome assembler for Illumina sequencing platform	<a href="http://bioinf.spbau.ru/spades">http://bioinf.spbau.ru/spades</a>	Bankevich <i>et al.</i> 2012
FastQC	Quality control of raw sequence data	<a href="http://www.bioinformatics.babraham.ac.uk/projects/fastqc/">http://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>	Andrews 2010
Flexbar	Sequence trimming and adapter removal	<a href="http://sourceforge.net/projects/flexbar/">http://sourceforge.net/projects/flexbar/</a>	Dodt <i>et al.</i> 2012
Artemis Release 8.0	Genome browser and annotation tool	<a href="http://www.sanger.ac.uk/science/tools/artemis">http://www.sanger.ac.uk/science/tools/artemis</a>	Rutherford <i>et al.</i> 2000
DNA Plotter	Genome visualization	<a href="http://www.sanger.ac.uk/Software/Artemis/circular/">http://www.sanger.ac.uk/Software/Artemis/circular/</a>	Carver <i>et al.</i> 2009
Basic Local Alignment Search tool (BLAST)	Comparison of sequence similarity	<a href="http://blast.ncbi.nlm.nih.gov/Blast.cgi">http://blast.ncbi.nlm.nih.gov/Blast.cgi</a>	Altschul <i>et al.</i> 1990
GAMOLA	Gene calling and automated annotation	<a href="http://fsweb2.schaub.ncsu.edu/TRKwebsite/index.htm">http://fsweb2.schaub.ncsu.edu/TRKwebsite/index.htm</a>	Altermann and Klaenhammer 2003
Staden package	Primer design and gap closure	<a href="http://staden.sourceforge.net/">http://staden.sourceforge.net/</a>	Staden <i>et al.</i> 2000
Vector NTI Advanced 11.5.2	DNA analysis and design suite	<a href="https://www.thermofisher.com/nz/en/home/life-science/cloning/vector-nti-software.html">https://www.thermofisher.com/nz/en/home/life-science/cloning/vector-nti-software.html</a>	Lu and Moriyama 2004
Geneious 6.1.15	DNA and protein analysis suite	<a href="http://www.geneious.com/">http://www.geneious.com/</a>	Kearse <i>et al.</i> 2012
tRNAscan-SE	tRNA prediction tool	<a href="http://lowelab.ucsc.edu/tRNAscan-SE/">http://lowelab.ucsc.edu/tRNAscan-SE/</a>	Lowe and Eddy 1997
Rfam 12.0	Database of RNA families	<a href="http://rfam.xfam.org/">http://rfam.xfam.org/</a>	Nawrocki <i>et al.</i> 2015
Pfam 29.0	Database of protein families	<a href="http://pfam.xfam.org/">http://pfam.xfam.org/</a>	Finn <i>et al.</i> 2016
Infernal 1.1	RNA prediction tool	<a href="http://infernal.janelia.org/">http://infernal.janelia.org/</a>	Nawrocki and Eddy 2013
MEGA6	Sequence alignment and phylogeny analysis	<a href="http://www.megasoftware.net/">http://www.megasoftware.net/</a>	Tamura <i>et al.</i> 2013
SignalP 4.1	Signal peptide prediction	<a href="http://www.cbs.dtu.dk/services/SignalP/">http://www.cbs.dtu.dk/services/SignalP/</a>	Petersen <i>et al.</i> 2011
LIPO	Lipo-box prediction	<a href="http://services.cbu.uib.no/tools/lipo">http://services.cbu.uib.no/tools/lipo</a>	Berven <i>et al.</i> 2006
CW-PRED	LPxTG motif prediction	<a href="http://bioinformatics.biol.uoa.gr/CW-PRED/">http://bioinformatics.biol.uoa.gr/CW-PRED/</a>	Fimereli <i>et al.</i> 2012
Pfam	Pfam prediction	<a href="http://pfam.xfam.org/">http://pfam.xfam.org/</a>	Finn <i>et al.</i> 2016
TETRA	Tetranucleotide usage pattern analysis	<a href="http://www.megx.net/tetra">http://www.megx.net/tetra</a>	Teeling <i>et al.</i> 2004
SPAAN	Adhesin prediction	<a href="http://sourceforge.net/projects/adhesin/files/SPAAN/">http://sourceforge.net/projects/adhesin/files/SPAAN/</a>	Sachdeva <i>et al.</i> 2005
PSORTb 3.0	Protein subcellular location prediction	<a href="http://www.psort.org/psortb/">http://www.psort.org/psortb/</a>	Yu <i>et al.</i> 2010
SubLoc	Protein subcellular location prediction	<a href="http://www.bioinfo.tsinghua.edu.cn/SubLoc/">http://www.bioinfo.tsinghua.edu.cn/SubLoc/</a>	Chen <i>et al.</i> 2006
LocTree3	Protein subcellular location prediction	<a href="https://roslab.org/services/loctree3/">https://roslab.org/services/loctree3/</a>	Goldberg <i>et al.</i> 2014
ISfinder	Insertion sequence element prediction	<a href="https://www-is.biotoul.fr/">https://www-is.biotoul.fr/</a>	Siguier <i>et al.</i> 2006

Alien_Hunter	Horizontal transferred region prediction	<a href="http://www.sanger.ac.uk/science/tools/alien-hunter">http://www.sanger.ac.uk/science/tools/alien-hunter</a>	Vernikos and Parkhill 2006
CAIcal	Codon adaption index calculation	<a href="http://genomes.urv.cat/CAIcal/">http://genomes.urv.cat/CAIcal/</a>	Puigbo <i>et al.</i> 2008
PKS/NRPS Analysis Web-site	Identification and domain prediction of NRPS/PKS	<a href="http://nrps.igs.umaryland.edu/">http://nrps.igs.umaryland.edu/</a>	Bachmann and Ravel 2009
NRPSSp	NRPS substrate predictor	<a href="http://www.nrpssp.com/execute.php">http://www.nrpssp.com/execute.php</a>	Prieto <i>et al.</i> 2012
REPFIND	Repeat identification	<a href="http://zlab.bu.edu/repfind/">http://zlab.bu.edu/repfind/</a>	Betley <i>et al.</i> 2002
CMG-biotools	Comparative genomics tools	<a href="http://www.cbs.dtu.dk/~dave/CMGtools/">http://www.cbs.dtu.dk/~dave/CMGtools/</a>	Vesth <i>et al.</i> 2013
MestRe Nova 11.0	H-NMR data analysis	<a href="http://mestrelab.com/">http://mestrelab.com/</a>	Willcott 2009
Chenomx NMR suite 7.0	NMR analysis suite with compound library	<a href="http://www.chenomx.com/">http://www.chenomx.com/</a>	Weljie <i>et al.</i> 2006
HMDB	Human metabolome database	<a href="http://www.hmdb.ca/">http://www.hmdb.ca/</a>	Wishart <i>et al.</i> 2007
RMDB	Bovine rumen metabolome database	<a href="http://www.rumendb.ca/cgi-bin/browse.cgi">http://www.rumendb.ca/cgi-bin/browse.cgi</a>	Saleem <i>et al.</i> 2013
MetaboAnalyst 3.0	Statistical analysis of metabolomics data	<a href="http://www.metaboanalyst.ca/">http://www.metaboanalyst.ca/</a>	Xia <i>et al.</i> 2015
Rockhopper	Transcriptome analysis suite	<a href="http://cs.wellesley.edu/~btjaden/Rockhopper/">http://cs.wellesley.edu/~btjaden/Rockhopper/</a>	McClure <i>et al.</i> , 2013
EDGE-pro	Transcriptome analysis suite	<a href="http://ccb.jhu.edu/software/EDGE-pro/">http://ccb.jhu.edu/software/EDGE-pro/</a>	Magoc <i>et al.</i> , 2013
<b>R packages</b>			
mixOmics	Regularised canonical correlation analysis	<a href="https://cran.r-project.org/web/packages/mixOmics/index.html">https://cran.r-project.org/web/packages/mixOmics/index.html</a>	Dejean <i>et al.</i> 2011
Predictmeans	Analysis of variance and Kruskal-Wallis test	<a href="https://cran.r-project.org/web/packages/predictmeans/index.html">https://cran.r-project.org/web/packages/predictmeans/index.html</a>	Luo <i>et al.</i> 2014
PMCMR	Analysis of variance	<a href="https://cran.r-project.org/web/packages/PMCMR/index.html">https://cran.r-project.org/web/packages/PMCMR/index.html</a>	Pohlert 2014
Vegan	Multivariate analysis	<a href="https://cran.r-project.org/web/packages/vegan/index.html">https://cran.r-project.org/web/packages/vegan/index.html</a>	Oksanen <i>et al.</i> 2007
Ape	Multivariate analysis	<a href="https://cran.r-project.org/web/packages/apex/index.html">https://cran.r-project.org/web/packages/apex/index.html</a>	Paradis <i>et al.</i> 2008
Perm	Multivariate analysis	<a href="https://cran.r-project.org/web/packages/perm/index.html">https://cran.r-project.org/web/packages/perm/index.html</a>	Fay and Fay 2010
MASS	Multivariate analysis	<a href="https://cran.r-project.org/web/packages/MASS/index.html">https://cran.r-project.org/web/packages/MASS/index.html</a>	Ripley <i>et al.</i> 2015
FactoMineR	Correspondence analysis	<a href="https://cran.r-project.org/web/packages/FactoMineR/index.html">https://cran.r-project.org/web/packages/FactoMineR/index.html</a>	Lê <i>et al.</i> 2008
doBy	Correspondence analysis	<a href="https://cran.r-project.org/web/packages/doBy/index.html">https://cran.r-project.org/web/packages/doBy/index.html</a>	Højsgaard <i>et al.</i> 2006
DESeq	Calculate differential gene expression	<a href="http://bioconductor.org/packages/release/bioc/html/DESeq.html">http://bioconductor.org/packages/release/bioc/html/DESeq.html</a>	Anders and Huber 2010



#### 2.2.21 Prediction of non-ribosomal peptide synthase (NRPS) domains and substrates

The identity of gene encoding NRPSs was predicted during the annotation. The adenylation domains are predicted by the PKS/NRPS Analysis website (Bachmann and Ravel 2009) (<http://nrps.igs.umaryland.edu/>). The substrates of adenylation domains were predicted by NRPSsp (Non-Ribosomal Peptide Synthase Substrate Predictor, <http://www.nrpsp.com/index.php>) (Prieto *et al.* 2012).

#### 2.2.22 Prediction of insertion sequence elements

Insertion sequences were predicted by IS Finder (Siguier *et al.* 2006) (<https://www-is.biotoul.fr/>) based on mobile elements from annotation, and REPFIND (Betley *et al.* 2002) (<http://zlab.bu.edu/repfind/>) was used to verify flanking repeat sequence utilised for mobilisation.

#### 2.2.23 Prediction of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)

The CRISPR elements were predicted by the CRISPR Finder (Grissa *et al.* 2007) (<http://crispr.u-psud.fr/Server/>). The number and sequence of direct repeats were identified, the spacer sequences were analysed by CRISPRTarget (Biswas *et al.* 2013) ([http://bioanalysis.otago.ac.nz/CRISPRTarget/crispr\\_analysis.html](http://bioanalysis.otago.ac.nz/CRISPRTarget/crispr_analysis.html)) to identify targets of inhibition.

#### 2.2.24 Metabolic pathway mapping

The metabolic pathway was constructed based on homology to experimentally validated genes and existing pathways in databases including the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000) and Metacyc (Caspi *et al.* 2014). The gene homology was compared based on BLOSUM62 (BLOcks SUBstitution Matrix) sequence alignment (Henikoff and Henikoff 1992) within Geneious (Kearse *et al.* 2012).

#### 2.2.25 Codon frequency, amino acid usage, BLAST matrix and Pan-core plot

Codon frequency, amino acid usage were calculated by CMG-biotools (Vesth *et al.* 2013). The genomes were compared by CMG-biotools to produce a BLAST matrix and a Pan-core plot based on a gene family criteria of 50% amino acid identity over 50% of protein length.

#### 2.2.26 Genome similarity analysis

Genome synteny was analysed using the programme PROmer (Delcher *et al.* 2003) in the the MUMmer sequence alignment package. The command used in this study was:

```
“./promer –prefix=outputname input01.fasta input02.fasta”
```

```
“./mummerplot outputname.delta”
```

The genomes were compared and visualised by BLAST Ring Image Generator (BRIG) to identify similar regions conserved between genomes (Alikhan *et al.* 2011).

#### 2.2.27 Gene homology analysis

The identification of homolog used experimentally validated genes as reference genes, and a BLASTP (Altschul *et al.* 1990) expectation value of  $e^{-05}$  was used as cutoff unless otherwise stated. The coverage and length of the query sequence was checked for truncation and gene disruption, and the surrounding genes were analysed to detect any potential operon structures between functionally connected genes. Conservation of operons between genomes were also analysed.

#### 2.2.28 Metabolomics

The proton nuclear magnetic resonance ( $^1\text{H}$  NMR) was used to distinguish the metabolites depleted by growth of methanogenic archaeon ISO4-H5. Uninoculated controls and samples prior and post growth were processed by filtration via Nanosep 3k Omega (PALL Corporation, Port Washington, NY, USA), with 3 kDa cut-off to remove protein molecules. 440  $\mu\text{L}$  of sample was mixed with 100  $\mu\text{L}$  of  $\text{NaHPO}_4$  buffer (300 mM, pH 7.4) and 60  $\mu\text{L}$  of 5 mM DSS in deuterated water ( $\text{D}_2\text{O}$ ) was used as internal standard.

The  $^1\text{H}$  NMR was conducted on a superconducting 700 MHz nuclear magnetic resonance spectrometer (Bruker Corporation, Billerica, MA, USA) at Massey University, Palmerston North, New Zealand. Each spectra has been adjusted individually with baseline correction via Whittaker Smoother algorithm and manual phase correction in MestReNova 11.0 (Willcott 2009) with the assistance of NMR specialist Dr. Linda Samuelson. Spectras were superimposed and auto-scaled, and aligned by reference peaks. The water peaks, imidazole peaks and reference peaks were removed from aligned spectra. Spectra regions containing peaks between chemical shift of 0.6 and 8.6 ppm were binned by 0.01 ppm via sum, and normalized by total average of 100%. The binned spectra were statistically analysed in Metaboanalyst 2.0 (Xia *et*

*al.* 2012), data was normalised by log transformation (corrected for heteroscedasticity, the unequal variability of variables) and scaled by Pareto scaling (uses the square root of the standard deviation as the scaling factor) (Brodsky *et al.* 2010). T-tests with *P*-value threshold of 0.05 were used to identify significantly depleted compounds in samples. The compounds were identified in Chenomx 7.0 library (Weljie *et al.* 2006). The identity of compound was verified by HSQC and TOCSY using the Human Metabolome Database as reference (Wishart *et al.* 2013).

#### 2.2.29 Transcriptome analysis of methanogenic archaeon ISO4-H5

##### 2.2.29.1 Experimental design

ISO4-H5 gene expression was investigated with two variables, the hydrogen levels and methyl-substrates. Nine different treatment groups were designed to include controls for each treatment (Table 5.03). Eight biological replicates were conducted in serum bottles, and the methane and hydrogen concentration in headspace samples was monitored daily. Four replicates were harvested at 120-hour and four replicates were harvested at 144-hour post inoculation. The samples were snap frozen in liquid N<sub>2</sub> and stored in -80 °C.

##### 2.2.29.2 RNA extraction

All benches, fumehood and equipment were cleansed with RNaseZap decontamination solution (Thermo Fisher Scientific) prior to each RNA extraction experiment. All procedures were performed on ice unless otherwise stated. The frozen sample was thawed in RNAprotect bacteria reagent (1 mL/g of sample, Qiagen) slowly over ice. The cell suspension was centrifuged (15,000 × *g*, 10 min, 4 °C), and the supernatant was decanted carefully. Each cell pellet was resuspended in 500 µL of RNA extraction buffer A, 210 µL of 20% SDS, 500 µL of phenol:chloroform:isoamyl alcohol (125:24:1, pH 4.5, Thermo Fisher Scientific) with 500 mg of RNase free Zirconia/Silica beads (0.1 mm:0.5 mm, 1:1). The cell suspension was lysed by bead beating twice (45s, 6.5 speed) in a Savant FastPrep FP120 (MP Biomedicals, Santa Ana, CA, USA) with 10 min intervals on ice to prevent samples from overheating. Alternatively, a Mini-Beadbeater (Biospec products INC. OK, USA) was used. Samples were incubated at 4 °C overnight. Samples were processed by bead beating three more times. Samples were centrifuged (10,000 × *g*, 5 min, 4 °C), and the top layer was transferred to RNase free microcentrifuge tubes. An equal volume of isopropanol and 10% volume of 5M ammonium acetate (pH 5.5) was mixed and stored at -20 °C for a minimum of one hour. The RNA was precipitated by centrifugation (10,000 × *g*, 30 min, 4°C). The isopropanol was

carefully removed, and the pellet was washed with 500  $\mu$ L of 70% RNase free ethanol (Analytical grade ethanol:Diethylpyrocarbonate (DEPC) treated water, 7:3), centrifuged ( $14,000 \times g$ , 5 min, 25  $^{\circ}$ C). The RNA/DNA pellet was dissolved in 50  $\mu$ L of nuclease free water. The quality of RNA/DNA was checked by gel electrophoresis using 2% agarose.

The tRNA and 5S rRNA was removed by MEGAclear Transcription Clean-Up Kit (Thermo Fisher Scientific) as instructed by the manufacturer. The DNA was then removed from samples by two consecutive DNase treatments; Baseline-ZERO DNase (Illumina Netherlands BV, Eindhoven, Netherlands) treatment was carried out as instructed by the manufacturer, and the TURBO DNA-free Kit (Thermo Fisher Scientific) was also used as the manufacturer instructed. DNA removal was validated by PCR. The RNA was quantified using a Qubit Fluorimeter (Thermo Fisher Scientific) and RNA quality was assessed using a 2100 Bioanalyser (Agilent Technologies, Santa Clara, CA, USA). Three  $\mu$ g of purified RNA was sent to BGI for rRNA depletion and sequenced using the Illumina HiSeq 2000 platform. Libraries with 200 bp inserts were generated and paired-end sequenced to produce 90 bp reads. The adapter sequence were removed by BGI.

#### 2.2.29.3 Sequence read alignment and normalisation

The quality of sequence reads in FASTQ files were assessed by FastQC (Andrews 2010), and reads below a Phred quality score of 28 were trimmed by Flexbar (Dodt *et al.* 2012). Rockhopper (McClure *et al.* 2013) is a programme specifically designed for RNA-seq data analysis. Reference based assembly was carried out with 2% allowed mismatches and default parameters. The reads were aligned to the reference genome based on Bowtie2 (Langmead and Salzberg 2012), where an index was created based on Burrows-Wheeler transform (Burrows and Wheeler 1994) for the reference genome, and reads were aligned exactly to the genome or aligned by seed regions of the read and extended with Smith-Waterman algorithm (Smith and Waterman 1981).

The read count from each sample was normalized by the upper quartile gene expression level excluding genes with zero expression, which was found to provide the most coherent result to the quantitative PCR data (Bullard *et al.* 2010). Rockhopper determines differential expression of each gene. The sum of normalized reads for two treatments were used to calculate differential expression based on the Anders and Huber approach (Anders and Huber 2010), using the statistical test for null hypothesis, negative binomial distributions was used as statistical model to compute two-sided *P*-value. *P*-values are influenced by number of

replicates, size of transcript and variance of transcripts across replicate samples. To reduce false positives,  $q$ -values were computed from the  $P$ -value based on the Benjamini-Hochberg correction (Benjamini 2010) with a false discovery rate of  $<1\%$ .

In addition to Rockhopper, a second software, EDGE-pro (Magoc *et al.*, 2013) was used to analyse the transcriptome reads, which follows similar procedure to Rockhopper. The reads were aligned using Bowtie2 (Langmead and Salzberg 2012) to the reference genome, where the raw coverage was converted to reads per kilobase of transcript per Million mapped reads (RPKM) by  $R = C \times L/r$ , where  $R$  is reads mapped to the gene,  $C$  is average coverage,  $L$  is length of gene,  $r$  is read length. The RPKM was then analysed by DESeq package (Anders and Huber 2010) in Rstudio, the differential gene expression was based on the relationship of mean to variance via binomial distribution. The results from EDGE-pro and DESeq was not used.

#### 2.2.29.4 Statistical determination of differential expression

Additional statistical testing was performed in order to select differentially expressed genes. Statistical tests were conducted using R studio 0.98.1049 (Rstudio Team, 2015). Three univariate tests were performed on the normalised reads; one-way analysis of variance (ANOVA), Kruskal-Wallis test (Kruskal and Wallis 1952) and Tukey's test (Tukey 1949) was performed to test the null hypothesis. The result of statistical tests would only be reliable when the data fits the assumption of each test. The two genes were considered differentially expressed between two treatments when both the  $q$ -value from Benjamini-Hochberg correction and the  $P$ -value from a reliable univariate test satisfies the criteria of  $<0.05$ .

The analyses were based on Bray-Curtis dissimilarity index using relative transcript abundance (Anderson *et al.* 2006), including non-metric multidimensional scaling (MDS), principal coordinate analysis (PCoA) (Legendre and Gallagher 2001), and permutation test for constrained analysis of principal coordinates (Legendre *et al.* 2011). The MDS is a multivariate statistical analysis used to reduce dimensionality and enable visualising similarities and dissimilarities between treatments (Oksanen *et al.* 2007). The PCoA is also used in dimension reduction and data visualization; it differs to MDS by preserving the covariance between data instead of distance (Gower and Legendre 1986). The group dispersion plot determines if one treatment is more variable than the others by permutation distribution under the null hypothesis (Anderson *et al.* 2006). Figures were produced in R studio with vegan, ape, perm and MASS package (Oksanen *et al.* 2007; Paradis *et al.* 2008; Fay and Fay 2010; Ripley *et al.* 2015).

The *R. flavefaciens* FD1 genome was re-annotated to correct for errors in open reading frame prediction and annotation. The re-annotation was carried out as described in Section 2.2.12.

Non-parametric multivariate analysis was performed to determine similarity of variation between different treatments. Bray-Curtis dissimilarity measure of distance was used to represent similarity (Bray and Curtis 1957). Multidimensional scaling (MDS) and principle coordinate analysis were applied to visualize the similarity and clustering of variables between treatments by reducing dimensions, where a low stress value indicates a reliable MDS. Multiple correspondence analysis were used to prioritise and visualise dissimilarity between gene sets presented as heatmaps and network diagrams. Relationships between gene expression of different organisms within the same treatment was assessed by regularised canonical correlation.

## Chapter 3

### Analysis of the methanogenic archaeon ISO4-H5 genome

#### 3.1 Introduction

Members of the order Methanomassiliicoccales, are methylotrophic methanogens and representatives of this order have been detected in various habitats, including landfills, rice fields, marine thermal vents, fresh water, and in the digestive tracts of termites, millipedes, chickens, ruminants and humans (Großkopf *et al.* 1998; Shinzato *et al.* 1999; Takai and Horikoshi 1999; Egert *et al.* 2003; Huang *et al.* 2003; Dridi *et al.* 2012; Horz *et al.* 2012; Iverson *et al.* 2012; Paul *et al.* 2012; Leahy *et al.* 2013; Padmanabha *et al.* 2013). The Methanomassiliicoccales are considered to be an important group of methanogens in the rumen environment and were originally referred to as Rumen Cluster C methanogens (Tajima *et al.* 2001; Baptiste *et al.* 2005). Their abundance in the rumen is highly variable, according to 16S ribosomal RNA gene surveys (Wright *et al.* 2006; Zhou *et al.* 2009; Huang *et al.* 2012), but on average, they constitute around 16% of the rumen archaeal community and are the second most abundant order of rumen methanogens (Janssen and Kirs 2008; Henderson *et al.* 2015). Representatives of these organisms have only recently been isolated in culture, and genomic information on members of the Methanomassiliicoccales are available for isolates from human, termite and bovine sources (Borrel *et al.* 2012; Gorlas *et al.* 2012; Borrel *et al.* 2013; Lang *et al.* 2015; NCBI Reference Sequence: NC\_020892.1, (Noel *et al.* 2016; Sollinger *et al.* 2016)).

An enrichment culture of a representative of the Methanomassiliicoccales was previously obtained from the rumen contents of a nine year old Romney wether sheep in NZ grazing a ryegrass-clover pasture diet (Jeyanathan 2010). This enrichment culture contained a methanogenic archaeon, designated ISO4-H5 and a Gram-negative bacterium, subsequently identified as being most closely related to *Succinivibrio dextrinosolvens*, and designated as strain H5. The methanogenic archaeon ISO4-H5 culture was enriched based on serial dilution and heat treatment with coenzyme M supplementation, without antibiotics (Jeyanathan 2010). The ISO4-H5 culture grew slowly and required three to four days to generate detectable methane in the culture headspace. The optical density of cultures after maximal CH<sub>4</sub> formation was very low and the ISO4-H5 cells could not be visualized via fluorescence microscopy at 420 nm due to an apparent lack of the fluorescent 8-hydroxy-5-deazaflavin cofactor, F<sub>420</sub> (Jeyanathan 2010). To gain an insight into the role of ISO4-H5 in methylotrophic methanogenesis in the rumen environment, its genome was sequenced. The ISO4-H5 genome

represents the first example from a member of the order Methanomassiliicoccales isolated from the ovine rumen.

## 3.2 Results

### 3.2.1 Genome sequencing results and assembly

*Genome assembly.* Whole genome sequencing yielded 1,025,715 reads and a total of 203,994,869 bp. The summary of the assembly statistics is displayed in Table 3.01. The assembly generated 29 scaffolds consisting of 690 contigs with an average size of 6,821 bp, the largest contig was 217,625 bp. The assembled genome of 29 scaffolds totals 4,660,809 bp, this provided 43.8× genome coverage for the combined genome data set containing ISO4-H5 and *S. dextrinosolvens* H5. There were 223 large contigs (> 500 bp) totalling 4,570,501 bp which were used for downstream analyses.

**Table 3.1 Genome assembly summary**

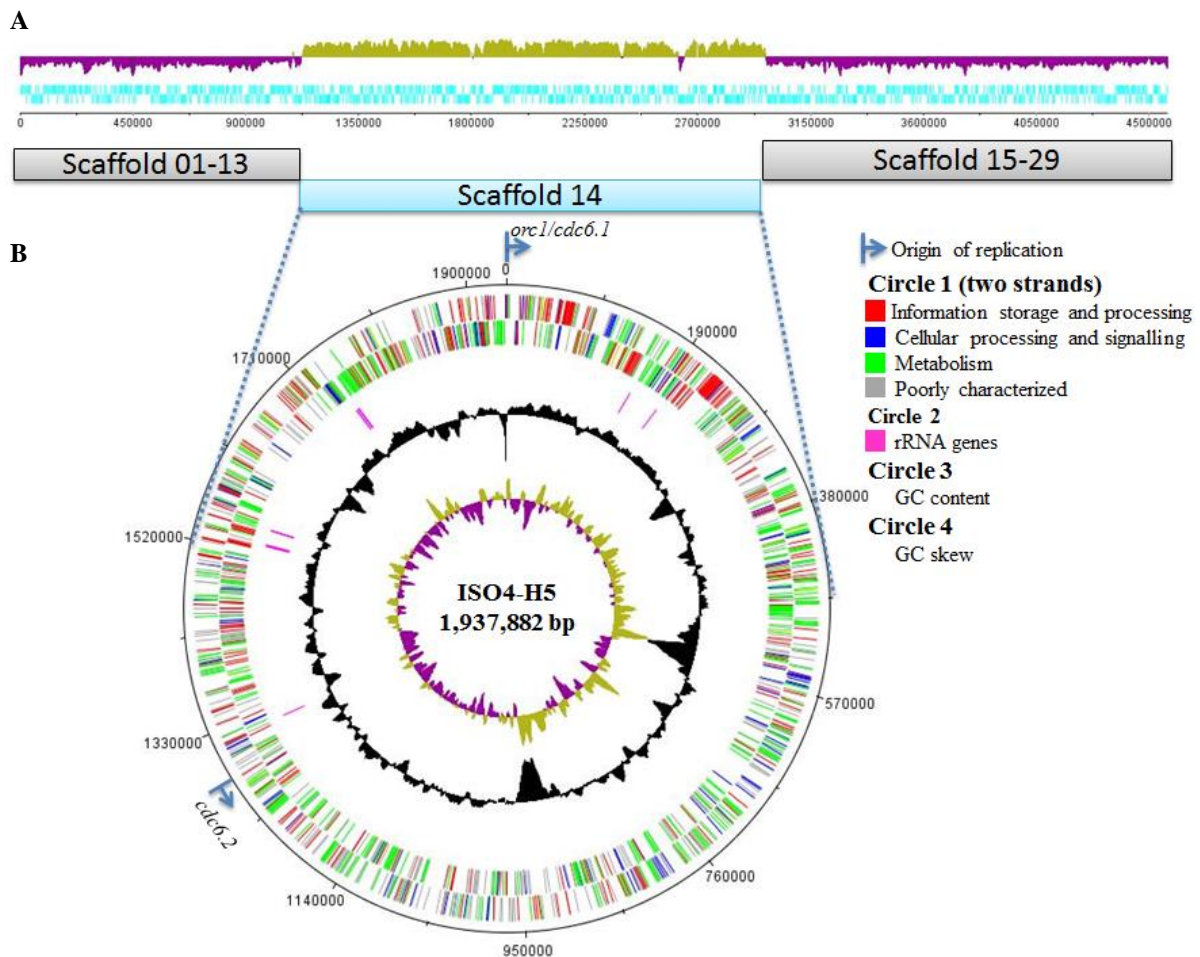
Genome sequence feature	Size or G+C%
<b>Enrichment metagenome</b>	
Total size	4,660,809 bp
Coverage	43.8×
Number of reads	1,025,715
Number of bases (bp)	203,994,879
Number of contigs	690
Number of scaffolds	29
Number of contigs >500 bp	223
G + C content	46.3%
G + C content omitting scaffold 14	40.71%
<b>Scaffold 14</b>	
G + C content	53.98%
Size	1,936,667 bp
Number of reads	356,637
Number of contigs	47
N50 contig size	73,897 bp

*ISO4-H5 scaffold assignment.* There were two distinct %G+C profiles observed in the enrichment genome (Figure 3.1A); scaffold 14 was 54%, while scaffolds 01 to 13 and 15 to 29 were 40.7%. The methanogen specific gene, methyl coenzyme-M reductase subunit A (*mcrA*) was identified within scaffold 14, and as a result, this scaffold was assigned to the methanogenic archaeon ISO4-H5. Scaffolds 01 to 13 and 15 to 29 were assigned to the *S. dextrinosolvens* H5 genome, and encompassed 176 contigs totaling 2,724,142 bp. A purified culture of *S. dextrinosolvens* H5 was included in the Hungate 1000 project, which aims to produce a reference set of rumen microbial genome sequences from cultivated rumen bacteria



and methanogenic archaea, together with representative cultures of rumen anaerobic fungi and ciliate protozoa (Creevey *et al.* 2014). The re-sequenced genome of H5 was reassembled into 106 contigs totalling 2,675,466 bp, and is predicted to encode 2,350 protein coding genes, (currently available through the DOE JGI web portal, Project ID 1026020). The *S. dextrinosolvens* H5 genome has not been analysed as part of this thesis.

**ISO4-H5 gap closure.** To complete the ISO4-H5 genome as a circular chromosome, it was necessary to close the gaps between the identified 47 contigs. The gap closure procedure is described in detail in Section 2.2.X. Nine rounds of gap closure were required to circularize the ISO4-H5 genome as a single contig. This involved a total of 163 PCRs and sequencing reactions to close gaps and to improve the quality of the genome sequence to ensure correct assembly and to resolve any remaining base conflicts.



**Figure 3.1 ISO4-H5 genome assembly and representation.** A. The %G+C content of the ISO4-H5 and *S. dextrinosolvens* H5 genome assembly. The %G+C content is displayed with a window size of 10,000 bp and a step size of 200 bp. The khaki area represents an above average % G+C content, and the purple area represents a below average % G+C content of the sequence data analysed. The cyan represents predicted ORFs in forward (upper) and reverse (lower) strand. B. Circular representation of the ISO4-H5 genome. Circles are numbered 1 (outermost) to 4 (innermost). Circle 1: predicted ORFs on the + and – strands respectively. ORFs are coloured based on the Clusters of Orthologous (COG) categories. Circle 2: location of the rRNA genes. Circle 3: % G+C

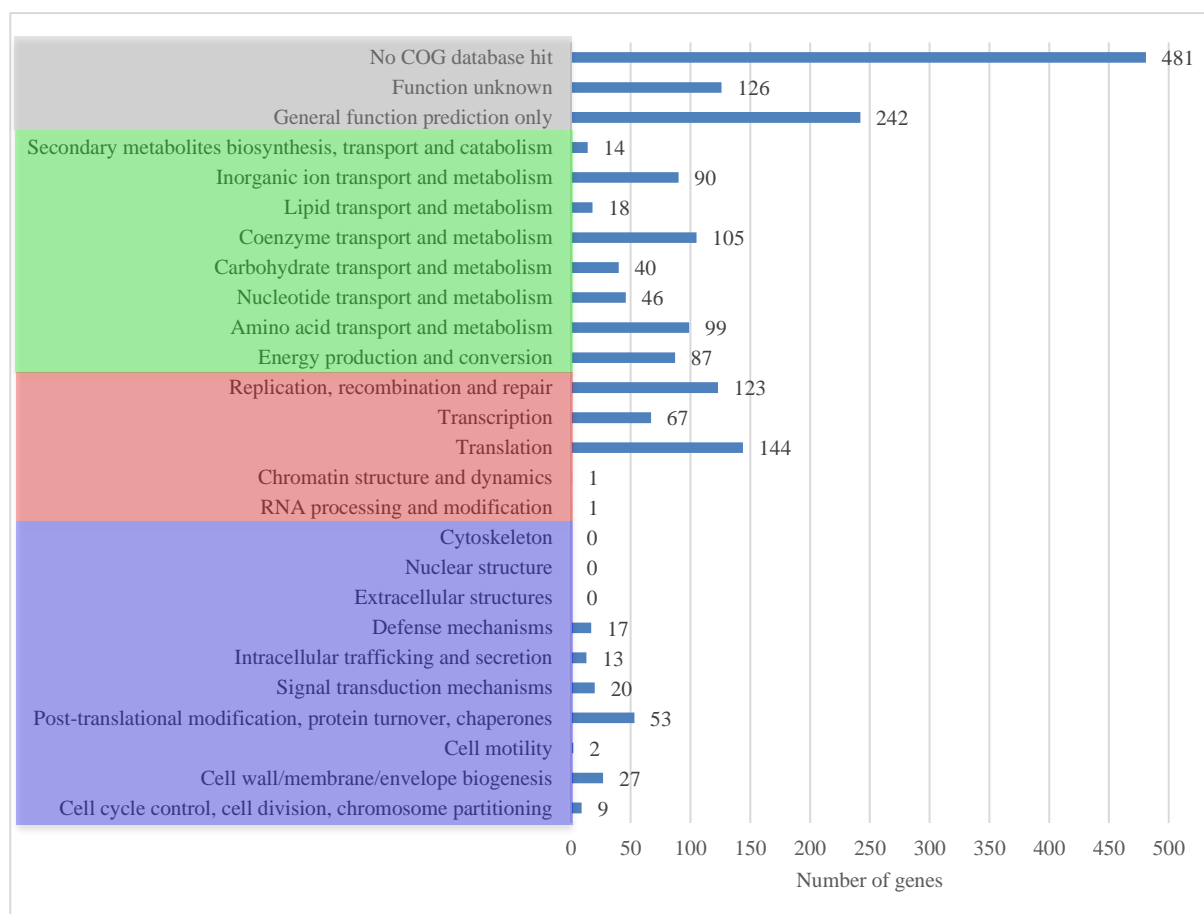
content. Circle 4: GC skew  $[(G-C)/(G+C)]$ , khaki coloured regions indicate values  $>1$ , purple  $<1$ . The identified origins of replication are represented as blue lines perpendicular to the outermost circle, the arrow indicates the 5' to 3' direction of DNA replication.

### 3.2.3 Genome properties

The general features of the ISO4-H5 genome are summarised in Table 3.2. Three pseudogenes were predicted in the genome of ISO4-H5, including a tRNA 2'-*O*-methylase (AR505\_0044), a site-specific recombinase (AR505\_1055) and a tryptophan synthase beta subunit *trpB* (AR505\_1151). No plasmids were identified in the genome of ISO4-H5 but a toxin/antitoxin system (AR505\_0857, 0858) possibly from a conjugative plasmid or prophage was identified.

**Table 3.2 General features of the ISO4-H5 genome**

Features	
Source	Ovine rumen
Project status	Complete
Genome size (bp)	1,937,882
%G+C content	54
Number of ORFs	1845
Coding area (%)	90
Contigs	1
rRNAs (5S, 16S, 23S)	3, 2, 2
tRNAs (with introns)	47 (3)
Non-coding RNAs	6
Insertion sequences	48
Prophage	1
CRISPR regions	1
Adhesin-like proteins	45



**Figure 3.2** The functional classification of the ISO4-H5 predicted genes based on the clusters of orthologous proteins (COG) database.

The functional categorizations of the predicted genes are summarized in Figure 3.2. Approximately one third of the total genes are annotated as hypothetical proteins with unknown function.

*Identification of the ISO4-H5 genome origin of replication.* The replication origin (*orc1/cdc6*) of the ISO4-H5 chromosome was identified by GC nucleotide skew  $[(G-C)/(G+C)]$  analysis (Figure 3.1B). Two *orc1/cdc6* genes were predicted in the genome (AR505\_0001, AR505\_1205). Two Origin Recognition Box (ORB) motifs were identified, 36 bp and 77 bp downstream of AR505\_0001, whereas AR505\_1205 gene had no ORB identified. Therefore, the *orc1/cdc6* gene (AR505\_0001) was predicted to be the true origin of replication of ISO4-H5.

*RNA.* ISO4-H5 possesses three copies of the 5S rRNA gene, two copies of the 16S rRNA gene and 2 copies of the 23S rRNA gene, but no operon structure was observed for these identified rRNAs (Table 3.3). Six ncRNAs were predicted to be present in the genome of ISO4-H5 (Table 3.3).

**Table 3.3 Predicted rRNA and ncRNA genes of ISO4-H5**

Start	End	Strand	E-value	Identity
<b>rRNA</b>				
161344	161465	+	3.10E-13	5S rRNA
200076	199955	-	3.10E-13	5S rRNA
1314829	1314950	+	3.10E-13	5S rRNA
1742102	1740637	-	0	16S rRNA
1552272	1550806	-	0	16S rRNA
1530665	1530555	-	1.40E-14	Pseudoknot of the domain G(G12) of 23S rRNA
1736778	1736668	-	1.40E-14	Pseudoknot of the domain G(G12) of 23S rRNA
<b>ncRNA</b>				
1465321	1465619	+	4.90E-54	Archaeal RNase P
882514	882820	+	4.10E-37	Archaeal signal recognition particle RNA
1892922	1892750	-	9.20E-29	Group II catalytic intron D1-D4-3
908140	908295	+	8.20E-08	Group II catalytic intron D1-D4-7
908845	908921	+	3.80E-06	Group II catalytic intron
1892698	1892603	-	6.30E-06	Group II catalytic intron D1-D4-1

The genome of ISO4-H5 encodes 47 tRNAs, including tRNAs corresponding to all 21 amino acids. Three tRNAs carry introns, namely methionine, tryptophan and tyrosine (Table 3.4).

**Table 3.4 Predicted tRNAs in the ISO4-H5 genome**

tRNA	Number of	
	tRNAs	Introns
tRNA-Ala	3	0
tRNA-Arg	4	0
tRNA-Asn	1	0
tRNA-Asp	1	0
tRNA-Cys	1	0
tRNA-Gln	2	0
tRNA-Glu	2	0
tRNA-Gly	3	0
tRNA-His	1	0
tRNA-Ile	1	0
tRNA-Leu	5	0
tRNA-Lys	2	0
tRNA-Met	3	1
tRNA-Phe	1	0
tRNA-Pro	3	0
tRNA-Ser	4	0
tRNA-Thr	3	0
tRNA-Trp	1	1
tRNA-Tyr	1	1
tRNA-Val	3	0
tRNA-Pyl	1	0
Pseudo tRNAs	1	0
Total	47	3

*Codon and amino acid usage.* The codon usage and the base content at the third codon position were calculated for each predicted coding gene sequence (Table 3.5) and are summarised in Figure 3.3B, in addition to the codon adaptation index (CAI) (Table A.3.2). The predicted

codon usage of ISO4-H5 has a strong bias towards C at the third codon position (Figure 3.3A). The amino acid usage is summarised in Figure 3.3C.

*Pyrrolysine usage.* ISO4-H5 possesses a complete operon predicted to encode the genes required for the biosynthesis of pyrrolysine and for the aminoacylation of a transfer RNA to pyrrolysine (Figure 3.6). Table 3.6 lists the predicted pyrrolysine-containing genes. The likelihood of amber codon read-through and incorporation of pyrrolysine is based on homology via BLASTp. The length of peptide extension beyond amber codon is included to indicate the likelihood of a mis-folded protein product.

*Insertion sequence (IS) elements.* Insertion sequence elements were predicted as described in Section 2.2.22. 48 IS elements were identified in the genome of ISO4-H5, and these elements account for 2.6% of the coding DNA sequence. Three different types of insertion sequences are found within the ISO4-H5 genome; transposons, group II introns and phage integrases. The IS elements found in ISO4-H5 belong to families IS200/IS605, IS481, IS5, IS1634, IS4, IS110, ISL3, IS91, IS481 and ISKra4, with IS200/IS605 family predicted with the highest copy number (Table 3.7).

**Table 3.5 Codon usage of ISO4-H5**

Amino acid	Codon	% Total Codons	% encoded amino acid per codon	Amino acid	Codon	% Total Codons	% encoded amino acid per codon
Ser (S)	UCA	0.5	7.1	Ala (A)	GCA	2.0	25.7
	UCC	2.8	42.8		GCC	3.5	44
	UCG	1.0	15.5		GCG	1.5	18.6
	UCU	0.5	8.5		GCU	0.9	11.8
	AGC	1.3	19.4	O (Pyl)	UAG	0.0	100.0
	AGU	0.4	6.7	Trp (W)	UGG	1.0	100.0
Phe (F)	UUC	3.6	92.3	Pro (P)	CCA	0.2	5.3
	UUU	0.3	7.7		CCC	2.2	56.6
Thr (T)	ACA	0.6	10.4		CCG	0.8	20.1
	ACC	3.8	66.2		CCU	0.7	18.0
	ACG	0.7	12.4	His (H)	CAC	1.1	67.4
	ACU	0.6	10.9		CAU	0.6	32.6
Asn (N)	AAC	3.0	76.9	Asp (D)	GAC	4.0	61.6
	AAU	0.9	23.1		GAU	2.5	38.4
Lys (K)	AAA	1.7	29.3	Arg (R)	CGA	0.0	0.3
	AAG	4.2	70.7		CGC	1.2	23.3
Glu (E)	GAA	2.3	34.1		CGG	0.2	3.3
	GAG	4.4	65.9		CGU	0.7	14.6
Tyr (Y)	UAC	2.6	70.4		AGA	0.7	14.8
	UAU	1.1	29.6		AGG	2.2	43.6
Val (V)	GUA	1.0	13.1	Ile (I)	AUA	1.0	14.3
	GUC	3.5	47.5		AUC	5.0	73.3
	GUG	2.1	28.4		AUU	0.8	12.4
	GUU	0.8	11	Gly (G)	GGA	3.2	41.2
Gln (Q)	CAA	0.3	10.9		GGC	1.9	25.2
	CAG	2.1	89.1		GGG	0.9	12.0
Met (M)	AUG	3.2	100.0		GGU	1.7	21.6
Cys (C)	UGC	1.2	74.5	Stop codons			
	UGU	0.4	25.5	<i>ochre</i>	UAA	13.4	
Leu (L)	CUA	0.2	2.7	<i>amber</i>	UAG	1.4 (2.7)*	
	CUC	3.5	42.5	<i>opal</i>	UGA	84.1	
	CUG	2.8	34.4	Translation initiator			
	CUU	1.0	11.6	M	AUG	93.6	
	UUA	0.2	2.7	L	GUG	1.7	
	UUG	0.5	6.1	V	UUG	4.7	

\*the number in brackets indicates total in-frame UAG, without considering its use as a STOP codon or as a Pyl-encoding codon

Identifier: H5.gb.k.fna

TotalBases: 1746297

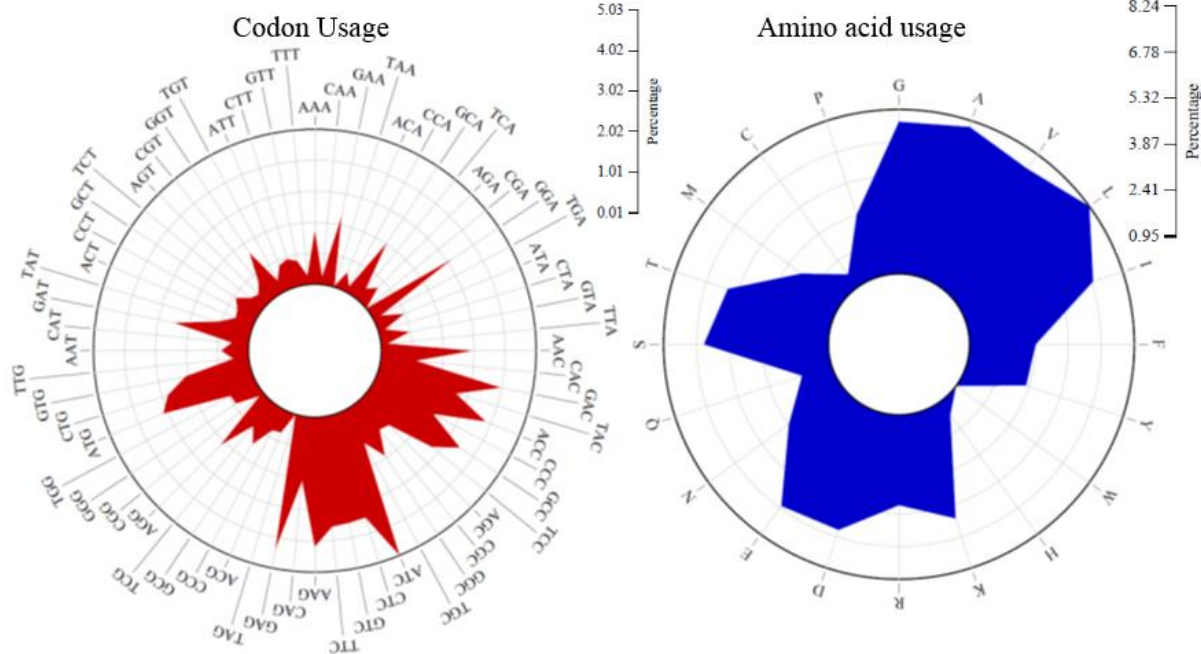
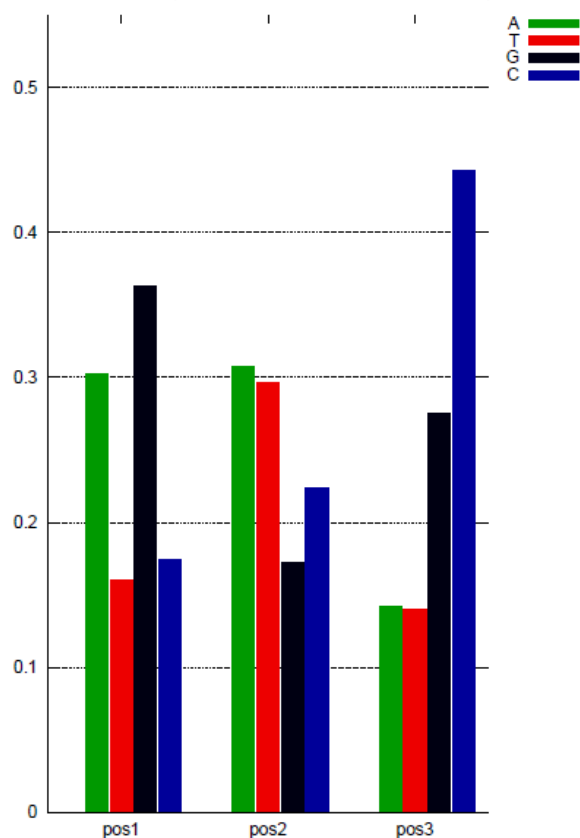
PerAT: 44.95

StDevAT: 0.05

Bias in third position: 0.4356

Bias in third position: 100% GC = +1 and 100% AT = -1

Position specific nucleotid usage



**Figure 3.3 The predicted codon and amino acid usage of ISO4-H5.** **A.** Base composition of codons in position one (pos1), two (pos2) and three (pos3). **B.** Codon usage percentage of ISO4-H5, scales are displayed on the right. **C.** Amino acid usage percentage of ISO4-H5, scales are displayed on the right.

**Table 3.6 Genes within the ISO4-H5 genome predicted to incorporate pyrrolysine in their encoded protein**

Locus tag	Gene name	Number of codons to next STOP*	Predicted class <sup>#</sup>
AR505_0107	Transposase	9	2
AR505_0316	Hypothetical protein	105	3
AR505_0319	Hypothetical protein	104	1
AR505_0344	Hypothetical protein	372	3
AR505_0479	Hypothetical protein	30	1
AR505_0499	2-methylcitrate synthase/Citrate synthase II PrpC	336	1
AR505_0516	Transposase IS605 OrfB family	34	2
AR505_0523	Transposase IS605 OrfB family	47	2
AR505_0544	Phosphoglycolate/pyridoxal phosphate phosphatase family protein	38	1
AR505_0560	Bifunctional phosphoglucose/phosphomannose isomerase	177	1
AR505_0572	Hypothetical protein	25	1
AR505_0573	Hypothetical protein	133	1
AR505_0580	Transposase IS605 OrfB family	51	2
AR505_0639	Transposase IS605 OrfB family	116	2
AR505_0669	Integrase catalytic subunit	9	2
AR505_0670	Hypothetical protein	14, 3	1
AR505_0708	Transposase	9	2
AR505_0772	Trimethylamine:corrinoide methyltransferase MttB	163	1
AR505_0816	Adhesin-like protein	319	1
AR505_0865	Hypothetical protein	84	3
AR505_0872	Hypothetical protein	60	1
AR505_0880	Hypothetical protein	219	1
AR505_0948	Hypothetical protein	266, 23	3
AR505_0952	Methanol corrinoide protein MtaC1	190	1
AR505_1022	Hypothetical protein	41	3
AR505_1089	CRISPR-associate endonuclease Cas3-HD	592	1
AR505_1141	Hypothetical protein	202, 93, 69	3
AR505_1203	Methanogenesis marker 8	72	1
AR505_1213	TPR repeat-containing protein	329	1
AR505_1215	Hypothetical protein	916	3
AR505_1265	Hypothetical protein	98	1
AR505_1289	Nitrogenase	237	1
AR505_1326	Hypothetical protein	5, 1	3
AR505_1327	Monomethylamine methyltransferase MtmB1	258	1
AR505_1328	Monomethylamine methyltransferase MtmB2	259	1
AR505_1332	Dimethylamine:corrinoide methyltransferase MtbB	112	1
AR505_1423	Transposase	9	2
AR505_1433	Geranylgeranyl reductase	158	1
AR505_1571	Hypothetical protein	50	3
AR505_1592	Hypothetical protein	63	3
AR505_1618	Geranylgeranyl reductase	71	1
AR505_1640	Phage integrase	9	2
AR505_1734	Hypothetical protein	57	3
AR505_1776	Group II intron-encoding maturase	80	1
AR505_1784	Adenylate kinase Adk	20	2
AR505_1787	Glycosyl transferase family protein	305	1

\*Distance between amber codon and next ochre/opal stop codon. <sup>#</sup>(Class 1) genes which have amber codon read-through and subsequent incorporation of pyrrolysine; (Class 2) genes that utilised the amber codon as a stop codon; (Class 3) genes with uncertain amber codon usage due to lack of homologous genes



**Table 3.7 IS elements of ISO4-H5**

Locus tag	Family	Source of most similar sequence
AR505_0213	IS110	<i>Shewanella oneidensis</i>
AR505_0525	IS110	<i>Haloferax volcanii</i>
AR505_1703	IS110	<i>Haloferax volcanii</i>
AR505_0372	IS1634	<i>Mycoplasma mycoides</i>
AR505_0528	IS1634	<i>Mycoplasma mycoides</i>
AR505_0574	IS1634	<i>Mycoplasma bovis</i>
AR505_1563	IS1634	<i>Mycoplasma mycoides</i>
AR505_0010	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0215	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0276	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0290	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0371	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0498	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0516*	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0523*	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0580*	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0639*	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0677	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0860	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0903	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0913	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_1219	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_1692	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_1765	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_1775	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_1777	IS200/IS605	<i>Methanosarcina mazei</i>
AR505_0440	IS4	<i>Lactobacillus reuteri</i>
AR505_1562	IS4	<i>Lactobacillus reuteri</i>
AR505_1714	IS4	<i>Lactobacillus reuteri</i>
AR505_0106	IS481	<i>Azotobacter vinelandii</i>
AR505_0107*	IS481	<i>Archaeoglobus fulgidus</i>
AR505_0708*	IS481	<i>Archaeoglobus fulgidus</i>
AR505_1423*	IS481	<i>Archaeoglobus fulgidus</i>
AR505_0669*	IS481	<i>Archaeoglobus fulgidus</i>
AR505_1640*	IS481	<i>Archaeoglobus fulgidus</i>
AR505_0147	IS5	<i>Methanosaeta thermophila</i>
AR505_0151	IS5	<i>Methanosaeta thermophila</i>
AR505_0784	IS5	<i>Methanosaeta thermophila</i>
AR505_1356	IS5	<i>Methanosaeta thermophila</i>
AR505_1458	IS5	<i>Methanosaeta thermophila</i>
AR505_0313	IS91	<i>Shewanella violacea</i>
AR505_0931	IS91	<i>Weeksella zoohelcum</i>
AR505_1543	IS91	<i>Azoarcus sp.</i>
AR505_1570	IS91	<i>Azoarcus sp.</i>
AR505_1697	IS91	<i>Azoarcus sp.</i>
AR505_1239	ISL3	<i>Acidithiobacillus caldus</i>
AR505_1574	ISL3	<i>Acidithiobacillus caldus</i>
<b>Group II intron maturase</b>		
AR505_1776*	ISKra4	<i>Legionella drancourtii</i>

\*Contains an in-frame amber codon

*CRISPR Elements.* ISO4-H5 contains several Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) related genes (AR505\_1089 – 1095) associated with a CRISPR region. The CRISPR element contains 35 repeats (bases 1153894 to 1155995), 29 bp in length. The consensus sequence of the repeat was GAGTTCCCCACGCATGTGGGGATGAACCG. The presence of the CRISPR associated proteins (AR505\_1090, AR505\_1091) suggested that the ISO4-H5 CRISPR/Cas system belongs to type I-E of the CRISPR-2 family. A total of 32 predicted spacer sequences with species-specific PAM (protospacer adjacent motif) were used to identify the potential targets of CRISPR RNAs (crRNAs). The predicted crRNA targets are displayed in Table 3.8, spacers with no database matches (spacers 1, 7, 8, 9, 10, 11, 13, 14, 16, 24, 26, 28, 29, 30, 31) were omitted.

**Table 3.8 CRISPR spacer homology**

Spacer Number	crRNA target*	Accession number	Score*
2	<i>Legionella pneumophila</i> str. Lorraine plasmid pLELO	NC_018141	18
3	KJ019151 <i>Synechococcus</i> phage ACG-2014f isolate Syn7803C24	KJ019146	18
4	<i>Burkholderia</i> sp. RPE64 plasmid p1	NC_021289	16
5	KF669658 <i>Acinetobacter</i> phage Presley	KF669658	16
6	<i>Halobacterium salinarum</i> R1 plasmid PHS2	NC_010369	16
12	<i>Silicibacter</i> sp. TM1040 mega plasmid	NC_008043	24
15	<i>Pantoea</i> sp. At-9b plasmid pPAT9B03	NC_014840	16
17	<i>Azospirillum brasilense</i> Sp245 plasmid AZOBR_p3	NC_016595	18
18	<i>Azospirillum lipoferum</i> 4B plasmid AZO_p2	NC_016586	18
19	JF974292 Cyanophage S-SSM2	JF974292	16
20	HM152763 <i>Mycobacterium</i> phage LeBron	HM152763	16
21	<i>Klebsiella pneumoniae</i> strain CAV1344 plasmid pCAV1344-250	NZ_CP011623	18
22	<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> strain 234-12 plasmid pKpn23412-362	NZ_CP011314	16
23	<i>Pantoea</i> sp. PSNIH1 plasmid pPSP-3a9	NC_CP010326	20
25	<i>Rahnella</i> sp. Y9602 plasmid pRAHAQ01	NC_015062	20
27	<i>Rhodococcus opacus</i> PD630 plasmid 1	NZ_CP003950	18
32	<i>Anabaena cylindrica</i> PCC 7122 plasmid pANACY.01	NC_019772	16

\*spacer sequence identified by BLAST based homology screening (Biswas *et al.* 2013). Score was calculated by matches (+1) and mismatches (-1) across the whole length of the spacer without gaps.

*Horizontal gene transfer.* There are 31 regions in the ISO4-H5 genome that are predicted to be horizontally acquired by Alien Hunter based on an atypical sequence composition (Table 3.9), and these regions account for 17.2% of the genome. A total of 243 genes are predicted across the 31 regions. There are three regions with a high likelihood of being horizontally transferred, regions 4, 7 and 15.

**Table 3.9 Predicted horizontal gene transfer regions of ISO4-H5**

Region	Base range	IVOM score*	% G+C	Number of genes	Average CAI*	Likelihood
1	1..10000	19.666	50.96	5	0.610	Low
2	62500..67500	19.99	59.69	1	0.891	Low
3	72500..77500	26.576	59.01	2	0.804	Low
4	305000..330000	22.246	47.01	32	0.611	High
5	397500..405000	24.731	47.13	6	0.557	Moderate
6	530000..545000	22.182	46.86	12	0.594	Moderate
7	550000..582500	54.567	41.46	30	0.528	High
8	592500..597500	16.055	47.77	4	0.603	Low
9	657500..670000	27.715	46.51	8	0.585	Moderate
10	687500..695000	20.556	56.3	1	0.757	Low
11	735000..742500	21.367	57.99	1	0.738	Low
12	755000..760000	16.127	48.09	6	0.6215	Low
13	785000..792500	14.62	60.78	1	0.716	Low
14	890000..897500	18.259	47.74	10	0.637	Low
15	907500..950000	52.2	42.13	35	0.546	High
16	1047500..1057500	18.524	56.1	1	0.715	Low
17	1147500..1157500	17.559	49.02	10	0.572	Low
18	1222500..1232500	18.971	52.59	2	0.679	Low
19	1285000..1290000	15.88	48.79	2	0.667	Low
20	1337500..1345000	18.998	48.17	8	0.650	Low
21	1452500..1457500	17.144	48.17	5	0.602	Low
22	1530000..1535000	16.147	52.57	2	0.723	Low
23	1557500..1562500	20.939	46.63	4	0.586	Moderate
24	1567500..1572500	19.414	58.17	4	0.8565	Low
25	1632500..1647500	29.172	46.22	11	0.562	Moderate
26	1670000..1682500	23.294	47.58	5	0.572	Moderate
27	1685000..1692500	14.707	49.38	9	0.612	Low
28	1735000..1747500	19.641	53.71	6	0.775	Low
29	1827500..1845000	21.204	48.25	9	0.615	Moderate
30	1852500..1857500	14.157	48.73	7	0.618	Low
31	1860000..1865000	29.168	45.37	4	0.555	Moderate

\*Atypical sequence composition is indicated by interpolated variable ordered motif (IVOM) score, average CAI (codon adaptation index) and % G+C. IVOM scores is annotation independent and it implemented variable order  $k$ -mers as reliable estimates of local sequence composition, as higher order motifs is more likely to capture deviation from genome background. A high regional IVOM score, abnormal % G+C and low average CAI is associated with a high likelihood of horizontal transfer.

Regions 7 and 15 have 42% G+C content in comparison to the average 54% G+C content of the genome. Both regions have atypical codon composition with low codon adaptation index (CAI), which indicates a bias in codon composition. More than half of the genes within these regions encode hypothetical proteins and the majority of these genes have top BLASTp hits originate from various non methanogenic micro-organisms. Transposases are also identified within, or adjacent to the regions. Additionally, region 7 contains an operon of putatively horizontally transferred genes, containing the dTDP-L-rhamnose biosynthesis genes *rfbA*, *rfbB*, *rfbC* (AR505\_0552 – 0554) with corresponding transporters downstream. Region 4 has a %G+C content of 47%, 7% below the average genome G+C content of 54%. However, much like region 7 and region 15, a high number of genes within the region are hypothetical proteins. A large cluster of hypothetical genes within this region and extended outside this region (AR505\_0333 to AR505\_0358) are only present 3' to 5'.

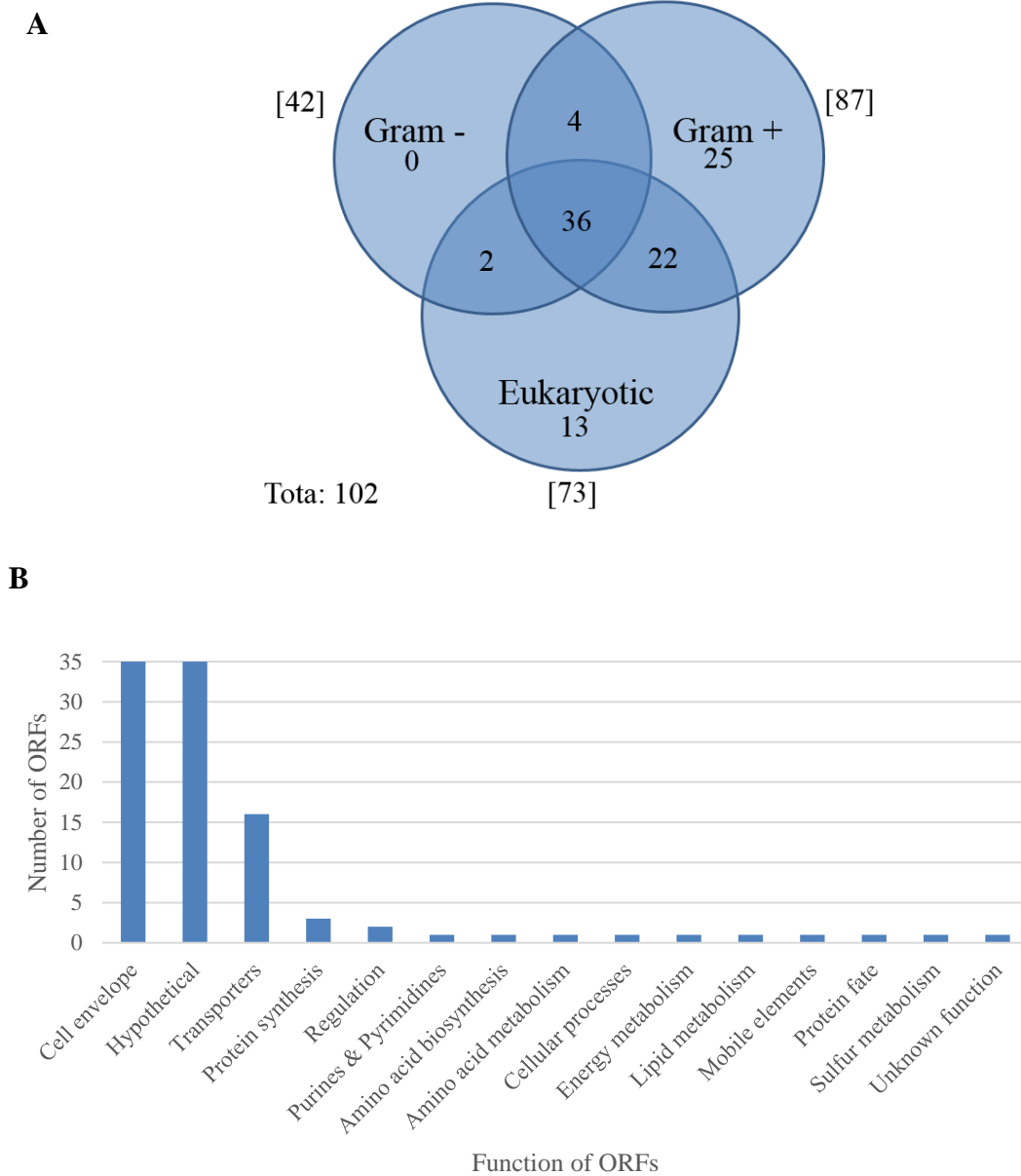
*Adhesin-like proteins.* There are 45 genes predicted to be adhesin-like proteins and these are displayed in Table 3.10, showing their predicted subcellular location, molecular weight, transmembrane helices, signal peptide and protein domains. Three adhesin-like protein are predicted to contain a quinonprotein alcohol dehydrogenase-like domain (AR505\_1521, 1549, 1707) and one adhesin-like protein is predicted to contain a pyrrolo-quinone beta-propeller repeat (AR505\_1524). In addition, there are seven adhesin-like proteins with molecular weights above 200 kDa (AR505\_0005, 0668, 0704, 0744, 0992, 1155, 1559).

**Table 3.10 Adhesin-like proteins**

Locus_tag	Size (kDa <sup>1</sup> )	TMH <sup>2</sup>	SignalP	SPAAN P <sub>ad</sub> -value	LOCTree3	PSORTb	SubLoc	Domain
AR505_0005	<b>209.81</b>	Yes	Yes	0.91	Secreted	Cellwall	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0061	197.01	Yes	Yes	0.96	Secreted	Cytoplasmic membrane	Extracellular	
AR505_0353	41.88	Yes	Yes	0.94	Secreted	Extracellular	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0354	24.54	Yes	Yes	0.81	Secreted	Unknown	Periplasmic	Listeria_Bacteriodes repeat (PF09479)
AR505_0355	33.34	No	Yes	0.86	Secreted	Extracellular	Extracellular	Putative Ig domain (PF05345)
AR505_0407	76.82	Yes	Yes	0.81	Secreted	Membrane or extracellular	Extracellular	Periplasmic binding protein (PF01497)
AR505_0594	73.45	Yes	Yes	0.81	Secreted	Extracellular	Extracellular	
AR505_0614	38.06	Yes	Yes	0.90	Plasma membrane	Membrane or extracellular	Extracellular	
AR505_0646	34.37	Yes	No	0.77	Secreted	Unknown	Extracellular	FKBP-type peptidyl-prolyl cis-trans isomerase (PF00254)
AR505_0654	24.21	Yes	Yes	0.86	Secreted	Extracellular	Extracellular	
AR505_0657	171.44	No	No	0.89	Secreted	Cellwall	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0658	22.21	Yes	Yes	0.81	Secreted	Membrane or extracellular	Periplasmic	
AR505_0660	103.94	Yes	Yes	0.86	Secreted	Membrane or extracellular	Extracellular	
AR505_0664	189.16	No	Yes	0.74	Secreted	Membrane or extracellular	Cytoplasmic	
AR505_0666	74.83	Yes	No	0.83	Secreted	Extracellular	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0668	<b>381.43</b>	Yes	Yes	0.98	Secreted	Cellwall	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0670	75.21	No	Yes	0.89	Secreted	Extracellular	Extracellular	
AR505_0704	<b>271.74</b>	Yes	Yes	0.83	Secreted	Cellwall	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0744	<b>382.06</b>	Yes	Yes	0.94	Secreted	Cellwall	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0807	74.08	Yes	Yes	0.84	Secreted	Unknown	Cytoplasmic	Listeria_Bacteriodes repeat (PF09479)
AR505_0851	45.14	Yes	Yes	0.77	Secreted	Cellwall	Cytoplasmic	Bacterial surface protein 26-residue repeat (TIGR02167)
AR505_0874	84.25	Yes	Yes	0.87	Secreted	Unknown	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0985	23.78	Yes	Yes	0.81	Secreted	Extracellular	Extracellular	
AR505_0989	24.16	Yes	Yes	0.84	Secreted	Extracellular	Extracellular	
AR505_0991	34.41	Yes	Yes	0.88	Secreted	Extracellular	Extracellular	
AR505_0992	<b>436.78</b>	Yes	Yes	0.97	Secreted	Cellwall	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_1032	49.03	Yes	No	0.82	Secreted	Extracellular	Extracellular	
AR505_1033	57.87	Yes	Yes	0.87	Secreted	Extracellular	Extracellular	Bacterial Ig-like domain (group 2) (PF02368)
AR505_1155	<b>294.67</b>	Yes	Yes	0.95	Secreted	Cytoplasmic membrane	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_1173	43.00	Yes	Yes	0.77	Secreted	Membrane or extracellular	Extracellular	
AR505_1290	146.37	Yes	Yes	0.83	Secreted	Membrane or extracellular	Extracellular	
AR505_1509	77.41	Yes	Yes	0.80	Secreted	Unknown	Extracellular	Bacterial Ig-like domain (group 2) (PF02368)
AR505_1521	95.56	Yes	Yes	0.70	Secreted	Membrane or extracellular	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_1524	66.24	Yes	Yes	0.75	Secreted	Extracellular	Extracellular	PQQ enzyme repeat (PF01011)
AR505_1532	95.53	Yes	Yes	0.76	Secreted	Membrane or extracellular	Periplasmic	
AR505_1534	91.42	No	Yes	0.87	Secreted	Membrane or extracellular	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_1547	39.47	Yes	Yes	0.74	Secreted	Membrane or extracellular	Periplasmic	
AR505_1549	65.82	Yes	Yes	0.77	Secreted	Extracellular	Extracellular	PQQ enzyme repeat (PF01011)
AR505_1559	<b>264.61</b>	Yes	Yes	0.84	Secreted	Cytoplasmic membrane	Cytoplasmic	Listeria_Bacteriodes repeat (PF09479)
AR505_1560	52.10	Yes	Yes	0.83	Secreted	Extracellular	Extracellular	Cohesin domain (PF00963)
AR505_1561	36.76	Yes	Yes	0.84	Secreted	Extracellular	Extracellular	
AR505_1707	76.48	Yes	Yes	0.70	Secreted	Membrane or extracellular	Periplasmic	PQQ enzyme repeat (PF01011)
AR505_1715	39.67	No	Yes	0.82	Secreted	Extracellular	Extracellular	
AR505_1741	41.34	Yes	Yes	0.74	Secreted	Extracellular	Extracellular	
AR505_1761	35.99	Yes	No	0.91	Secreted	Unknown	Extracellular	

<sup>1</sup> kDa: predicted molecular weight in thousand Daltons, proteins above 200 thousand Daltons are displayed in bold. <sup>2</sup> TMH: transmembrane helices.

*Secretome*. The ISO4-H5 genome encodes 101 ORFs predicted to contain a signal peptide (Figure 3.4A). Among the ORFeome predicted to be exported, most ORF functions have been classified under cell envelope, hypothetical or transporters (Figure 3.4B).

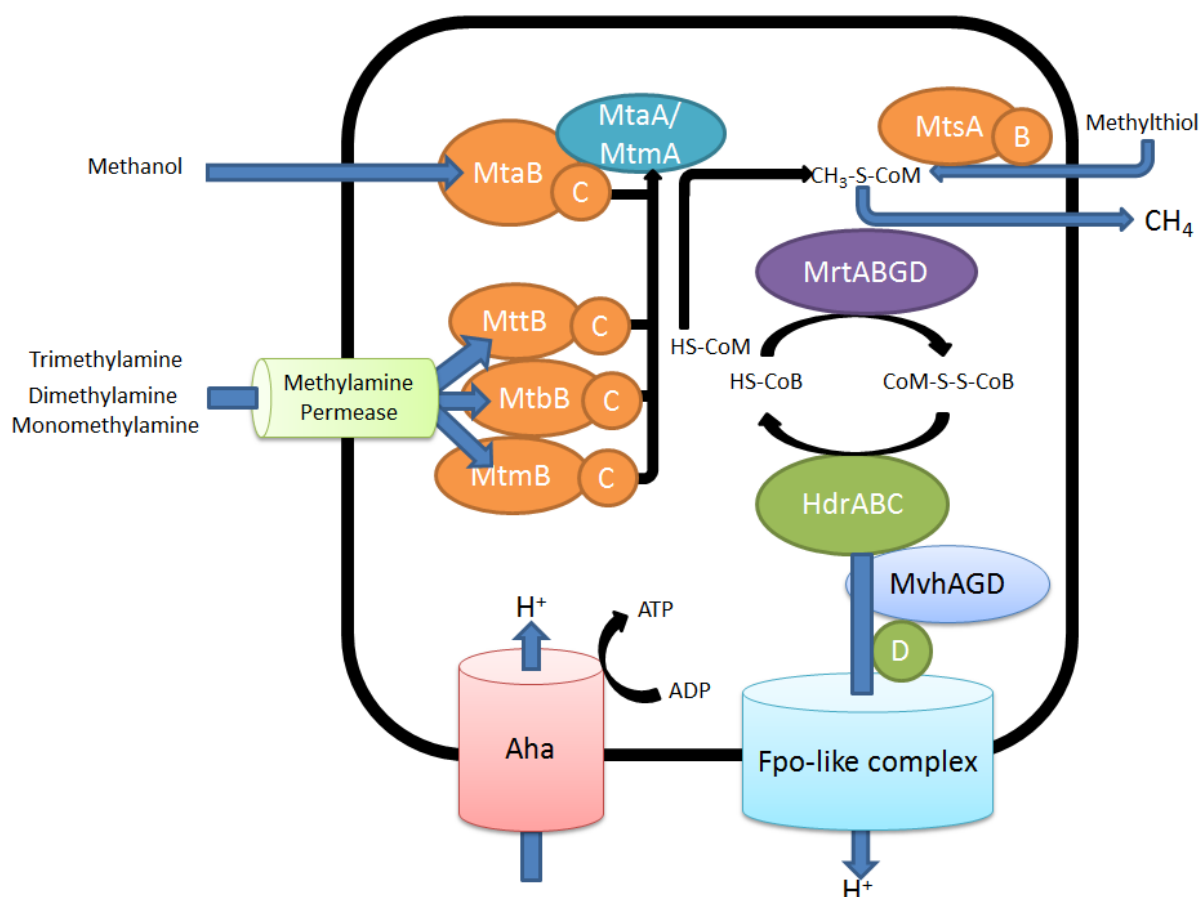


**Figure 3.4 Classification of ISO4-H5 signal peptides and their associated ORFs** **A.** Signal peptides classified according to SignalP training models. **B.** Functional classification of ISO4-H5 ORFs containing signal peptides

### 3.2.4 Metabolic pathway reconstruction

ISO4-H5 genes with a predicted metabolic function are listed in Table A.3.1. A total of 716 genes, representing 49% of the genome, have been assigned to a metabolic pathway.

#### *Energy metabolism*



**Figure 3.5 Methylotrophic methanogenesis of ISO4-H5** The enzymes involved are: methanol:corrinoid methyltransferase (MtaB), monomethylamine:corrinoid methyltransferase (MtmB), dimethylamine:corrinoid methyltransferase (Mtb), trimethylamine:corrinoid methyltransferase (MttB), methylcobalamin:coenzyme M methyltransferase (MtaA), methylthiol:coenzyme M methyltransferase (MtsA). The corresponding cognate corrinoid protein to each methyltransferase (C) with exception to methylthiol as (B). Methyl coenzyme M reductase (MrtABGD), heterodisulfide reductase (HdrABC), methyl viologen hydrogenase (MvhAGD). Components involved in membrane potential generation and energy generation are Fpo-like complex and ATP synthase (Aha).

*Methanogenesis.* Based on the genome sequence, methanogenesis is the sole energy generating mechanism of ISO4-H5. The genes predicted to be involved in methanogenesis are summarized in Figure 3.5. The ISO4-H5 genome encodes genes required for energy generation via methylotrophic methanogenesis, but lacks the genes encoding hydrogenotrophic methanogenesis prior to the methyl coenzyme reductase (Mrt) step. ISO4-H5 is strictly dependent on direct reduction of methylated compounds for methylotrophic methanogenesis.

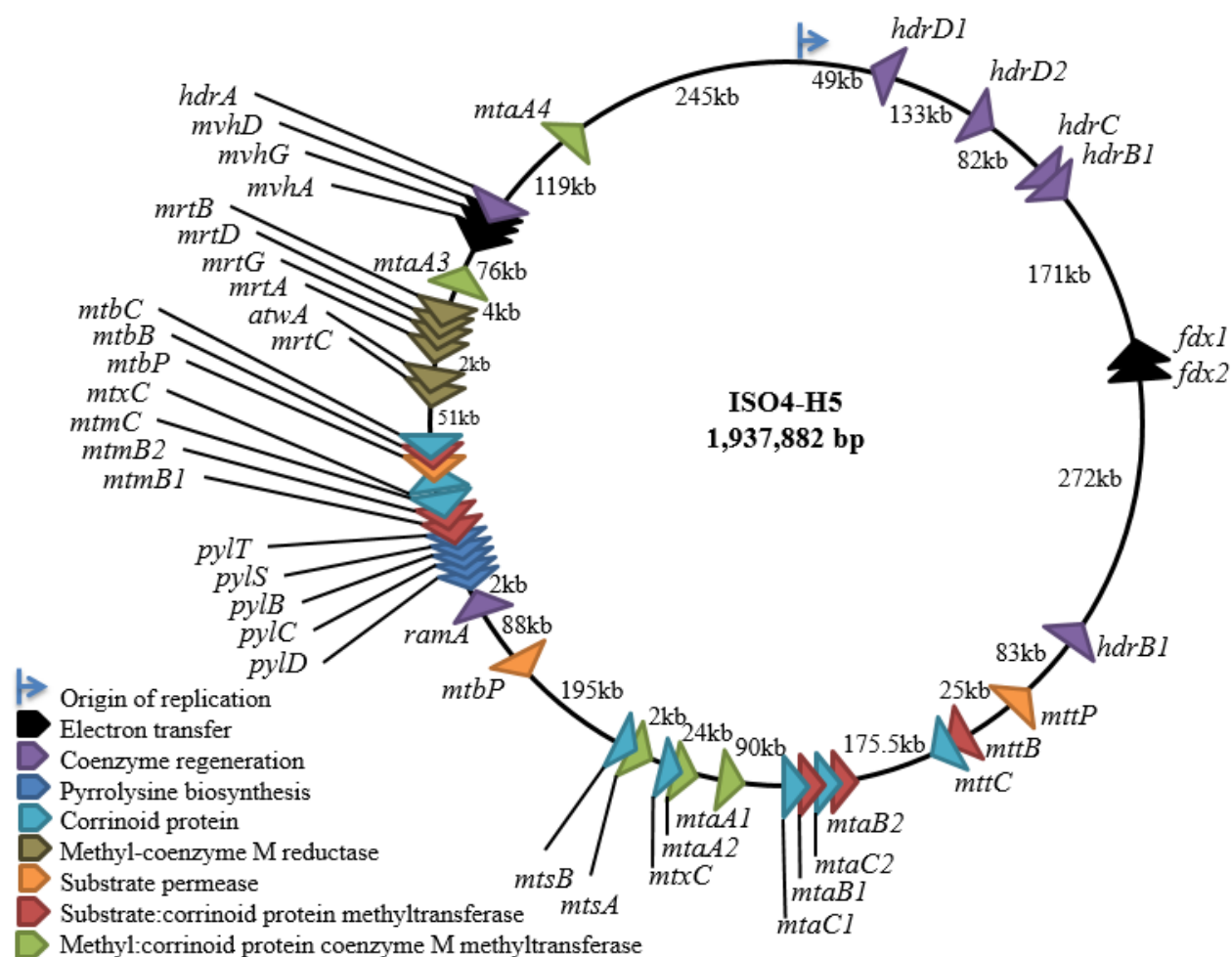
The ISO4-H5 genome encodes specific methyltransferases for utilising methylated compounds as substrates, including methanol, methylthiol, mono-, di-, and tri-methylamine (Figure 3.6). The *mta* operon required for methanol utilisation consists of two methanol:corrinoid methyltransferases, *mtaB1*, *mtaB2* (AR505\_0949, AR505\_0951), and two associated methanol corrinoid proteins, *mtaC1*, *mtaC2* (AR505\_0950, AR505\_0952) (Figure 3.6). The *mtaC2* (AR505\_0952) gene is predicted to require amber codon read-through.

The monomethylamine and dimethylamine utilisation genes are also present in operons, including two copies each of monomethylamine:corrinoid methyltransferase *mtmB* (AR505\_1327 and AR505\_1328), and the corresponding monomethylamine corrinoid protein *mtmC* (AR505\_1329 and AR505\_1330), in addition to a dimethylamine permease *mtbP* (AR505\_1331), a dimethylamine:corrinoid methyltransferase *mtbB* (AR505\_1332), and a dimethylamine corrinoid protein *mtbC* (AR505\_1333) (Figure 3.6). The trimethylamine utilisation genes, trimethylamine:corrinoid methyltransferase *mttB* and the corresponding trimethylamine corrinoid protein *mttC*, form a small operon (AR505\_0772, 0773) far upstream of the other methylamine utilisation genes (Figure 3.06). The pyrrolysine biosynthesis operon (AR505\_1322 – 1325) and *ramA* (AR505\_1320) were identified adjacent to the methylamine utilisation operon (Figure 3.5). Four homologues of methylcobalamin:coenzyme M methyltransferase *mtaA/mtmA* were also found in the ISO4-H5 genome.

ISO4-H5 possesses genes encoding for methylthiol utilisation. AR505\_1066 is highly homologous to methylthiol:coenzyme M methyltransferase from *Methanosarcina barkeri*, with 71.6% aa identity, and the corresponding corrinoid protein AR505\_1067 shares 50.2% amino acid identity to the *M. barkeri* methylthiol corrinoid protein. The methylthiol utilising operon is located adjacent to one of the methyltransferase cognate corrinoid genes (although with uncertain substrate specificity) and one of the *mtaA* genes (Figure 3.6).

The reduction of methyl-coenzyme M to methane is carried out by the methyl-coenzyme M reductase enzyme complex (*mrt*). ISO4-H5 harbors the gene organization of a MCRII/Mrt-type system, with four genes arranged in an operon structure, *mrtBDGA* (AR505\_1399 to AR505\_1396), with *mrtC* (AR505\_1391) and methyl-CoM reductase component A1 *atwA* (AR505\_1392) located downstream (Figure 3.6).





**Figure 3.6 Methanogenesis genes of ISO4-H5.** Genes predicted to be involved in methanogenesis in the ISO4-H5 genome are displayed with gene name followed by locus tag number. Heterodisulfide reductase (*hdr*), ferredoxin (*fdx*), trimethylamine permease (*mttP*), trimethylamine:corrinoid methyltransferase (*mttB*), trimethylamine corrinoid protein (*mttC*), methanol:corrinoid methyltransferase (*mtaB*), methanol corrinoid protein (*mtaC*), methyl:coenzyme M methyltransferase (*mtaA*), corrinoid protein with unknown substrate (*mtxC*), bifunctional methylthiol:corrinoid methyltransferase (*mtsA*), methylthiol corrinoid protein (*mtsB*), dimethylamine permease (*mtbP*), methylamine methyltransferase corrinoid activation protein (*ramA*), pyrrolysine biosynthesis protein (*pyl*), amber suppressor tRNA (*pylT*), monomethylamine:corrinoid methyltransferase (*mtmB*), monomethylamine corrinoid protein (*mtmC*), dimethylamine:corrinoid methyltransferase (*mtbB*), dimethylamine corrinoid protein (*mtbC*), methyl-coenzyme M reductase (*mrt*), methylviologen hydrogenase (*mvh*). Distance between genes are displayed in kb. Figure not drawn to scale.

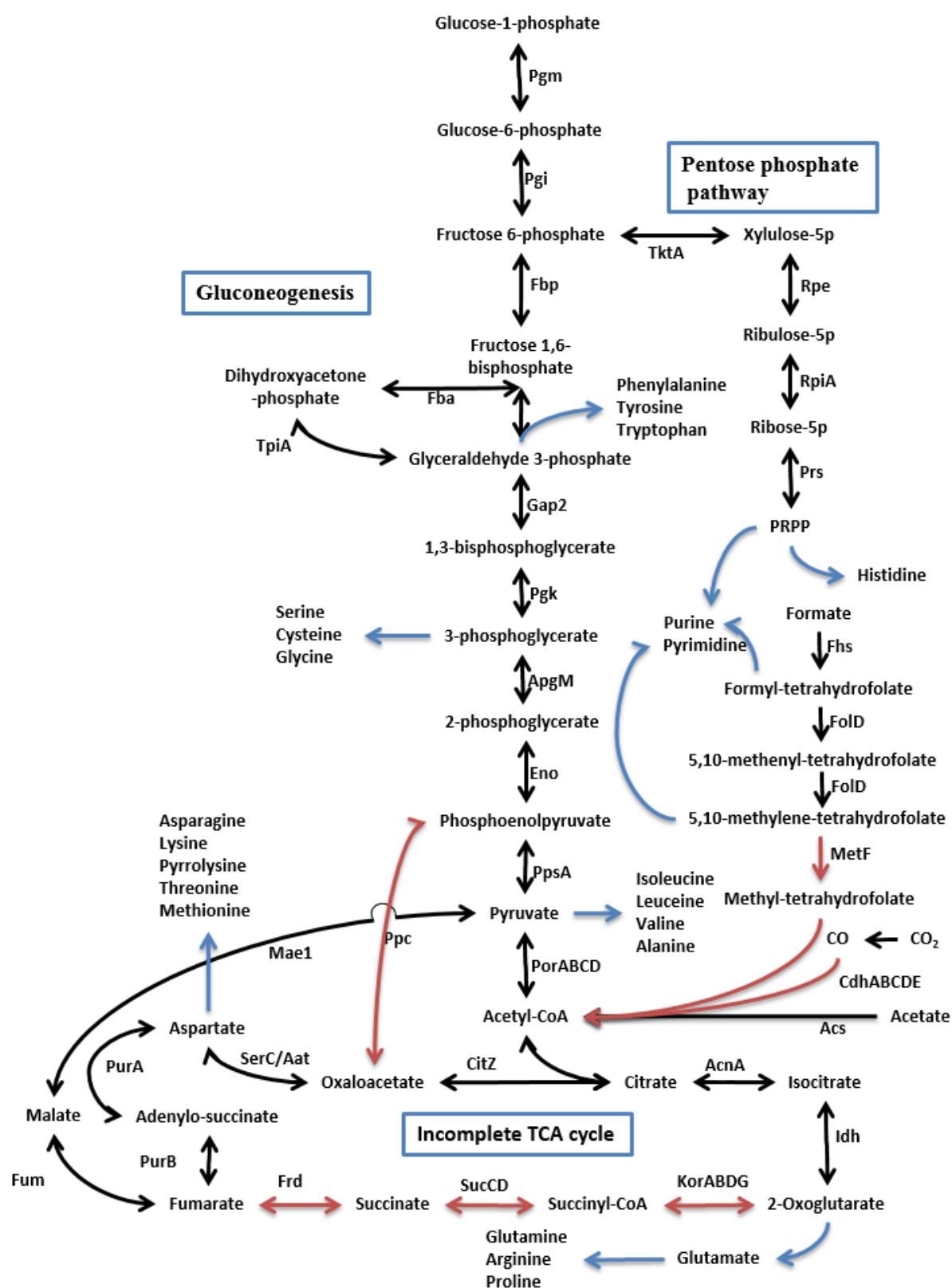
An alcohol dehydrogenase gene, *adhE* (AR505\_0483), homologous to an alcohol dehydrogenase identified in *Clostridium acetobutylicum* ATCC 824 and with a COG match to the alcohol dehydrogenase class IV Adh7 domain is also present in the ISO4-H5 genome.

**Methanogenesis cofactors.** During the last step of methanogenesis, coenzyme B (CoB) is used by the Mrt complex for methyl-coenzyme M reduction to produce methane (Figure 3.5). This process utilises the reducing potential carried by HS-CoB and generates CoM-S-S-CoB (heterodisulfide). Both cofactors need to be recycled into an active state, and this is carried out

by a heterodisulfide reductase (Hdr). ISO4-H5 encodes four Hdr subunits (HdrABCD), with *hdrB* and *hdrC* located in an operon (AR505\_0273, AR505\_0274), while *hdrA* (AR505\_1479) and a second copy of *hdrB* (AR505\_0679) are located elsewhere on the genome (Figure 3.6). There are also two copies of *hdrD* (AR505\_0040, AR505\_0168) in the genome without the usual partner gene *hdrE*. ISO4-H5 encodes the methyl viologen hydrogenase MvhAGD required for coupling H<sub>2</sub> to provide reducing potential for the reducing CoM-S-S-CoB and ferredoxin via electron bifurcation (*mvhADG*, AR505\_1476 – 1478) (Figure 3.6).

*Generation of membrane potential from methanogenesis.* The protons produced by regeneration of cofactors involved in methanogenesis is predicted to be removed from the cell by a proton pump, which generates a membrane potential to drive the ATP synthase. Two operons consisting of 11 genes encoding F<sub>420</sub>H<sub>2</sub> dehydrogenase (Fpo) have been found in ISO4-H5 (*fpoABCDHIJKLMN*, include genes, AR505\_1622-1625, AR505\_1627-1633). Two subunits, *fpoF* and *fpoO*, found in other methanogens are not encoded in the ISO4-H5 genome. Homologues of the nine A<sub>1</sub>A<sub>0</sub> ATP synthase subunits are present (*ahaHIKECFABD*, include genes, AR505\_1818- 1826) and is predicted to utilise the membrane potential generated by the Fpo-like complex to produce ATP.

*Methanogenesis marker proteins.* Homologues of methanogenesis marker proteins 1, 2, 3, 4, 5, 6, 7, 8, 11, 13, 15, 16, and 17 are present within the ISO4-H5 genome, while marker proteins 9, 10, 12 and 14 are absent.



**Figure 3.7 ISO4-H5 central carbon metabolism.** Black arrow represents functional pathways, red arrow represents pathways absent in ISO4-H5 but present in other members of Methanomassiliicoccales, blue arrows represents connections to other biosynthetic pathways. Formate-tetrahydrofolate ligase (Fhs), NADP-dependent methylene tetrahydrofolate dehydrogenase (FolD), 5,10-methylenetetrahydrofolate reductase (MetF), carbon-

monoxide dehydrogenase complex (CdhABCDE), acetyl CoA synthetase (Acs), phosphoglucomutase (Pgm), phosphoglucose isomerase (Pgi), fructose-1,6-bisphosphatase (Fbp), fructose-bisphosphate aldolase (Fba), glyceraldehyde-3-phosphate dehydrogenase (Gap2), phosphoglycerate kinase (Pkg), phosphoglycerate mutase (ApgM), phosphopyruvate hydratase (Eno), phosphoenolpyruvate synthase (PpsA), pyruvate:ferredoxin oxidoreductase (PorABCD), citrate synthase (CitZ), aconitate hydratase (AcnA), isocitrate dehydrogenase (Idh), 2-oxoglutarate synthase (KorABDG), succinyl-CoA synthetase (SucCD), fumarate reductase (Frd), fumarate hydratase (Fum), adenylosuccinate lyase (PurB), adenylosuccinate synthetase (PurA), phosphoserine aminotransferase/aspartate aminotransferase (SerC/Aat), malate dehydrogenase (Mae1), phosphoenolpyruvate carboxylase (Ppc), transketolase (TktA), ribulose-phosphate 3-epimerase (Rpe), ribose-5-phosphate isomerase (RpiA), ribose-phosphate diphosphokinase (Prs), phosphoribosyl pyrophosphate (PRPP).

ISO4-H5 is predicted to encode the genes required for a partial TCA cycle, gluconeogenesis and the pentose phosphate pathway.

*One-carbon metabolism.* ISO4-H5 has an incomplete 1-carbon metabolism and is incapable of producing acetyl-CoA *de novo* from carbon monoxide and formate, because it does not encode methyltetrahydrofolate:corrinoid methyltransferase, 5,10-methylenetetrahydrofolate reductase or the carbon-monoxide dehydrogenase complex (*cdhABCDE*) (Figure 3.7, Table 3.11). ISO4-H5 has *fold* encoding the bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/5,10-methylene-tetrahydrofolate cyclohydrolase (AR505\_1639) used to convert N<sup>10</sup>-methenyl-tetrahydrofolate to 5,10-methylene-tetrahydrofolate or N<sup>10</sup>-formyl-tetrahydrofolate.

ISO4-H5 encodes an acetyl-CoA synthetase, *acs*, as well as a holoenzyme of pyruvate synthase *porABCD* (Table 3.11). Acetate is predicted to be the carbon source of ISO4-H5. The acetyl-CoA synthetase appears to be a single, fused enzyme instead of the usual two subunits found in other organisms. The incorporation of acetyl-CoA into the TCA cycle produces pyruvate, which proceeds towards gluconeogenesis and subsequently the pentose phosphate pathway.

**Table 3.11 Genes involved in central carbon metabolism**

Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimental homologue	validated
AR505_1217	Formate--tetrahydrofolate ligase	<i>fhs</i>	52.1	Rankin <i>et al.</i> , 1993	
AR505_1639	Bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase	<i>folD</i>	44.4	Shen <i>et al.</i> , 1999	
AR505_0076	Thymidylate synthase	<i>thyX</i>	27.8	Myllykallio <i>et al.</i> , 2002	
AR505_1282	Acetyl-CoA synthetase (ADP-forming)	<i>acs</i>	36.4, 46.0*	Musfeldt <i>et al.</i> , 1999	
AR505_0431	Pyruvate synthase subunit $\alpha$	<i>porA</i>	39.9	Smith <i>et al.</i> , 1997	
AR505_0432	Pyruvate synthase subunit $\beta$	<i>porB</i>	43.0	Smith <i>et al.</i> , 1997	
AR505_0429	Pyruvate synthase subunit $\gamma$	<i>porC</i>	45.1	Smith <i>et al.</i> , 1997	
AR505_0430	Pyruvate synthase subunit $\delta$	<i>porD</i>	37.6	Smith <i>et al.</i> , 1997	
AR505_0022	Fumarate hydratase	<i>fumA</i>	43.5	Shieh & Whitman, 1987	
AR505_0023	Fumarate hydratase	<i>fumB</i>	45.3	Shieh & Whitman, 1987	
AR505_1780	Malate dehydrogenase	<i>maeI</i>	12.3	Shieh & Whitman, 1987, Shieh & Whitman, 1987	
AR505_0742	Adenylosuccinate lyase	<i>purB</i>	25.6	Zhang <i>et al.</i> , 2008	
AR505_1168	Adenylosuccinate synthetase	<i>purA</i>	44.1	Zhang <i>et al.</i> , 2008	
AR505_0499,	Citrate synthase	<i>citZ1</i> ,	28.2,	Jongsareejit <i>et al.</i> , 1997	
AR505_0678		<i>citZ2</i>	9.8	Khomyakova <i>et al.</i> , 2011	
AR505_0592	Aconitate hydratase	<i>acnA</i>	51.2	Gruer & Guest, 1994	
AR505_0531	Isopropylmalate/isocitrate dehydrogenase	<i>idh/leuB1</i>	22.3	Pitson <i>et al.</i> , 1999	
AR505_1608	Phosphoglucumutase	<i>pgm</i>	35.5	Jansen <i>et al.</i> , 1982	
AR505_0560,	Bifunctional	<i>pgi1</i> , <i>pgi2</i>	31.3, 31.6	Hansen <i>et al.</i> , 2004	
AR505_1769	phosphoglucose/phosphomannose isomerase				
AR505_1149	Fructose-1,6-bisphosphatase	<i>fbp</i>	50.1	Jansen <i>et al.</i> , 1982	
AR505_0932	Fructose-bisphosphate aldolase	<i>fba</i>	55.5	Jansen <i>et al.</i> , 1982	
AR505_0154	Glyceraldehyde-3-phosphate dehydrogenase	<i>gap2</i>	49.3	Jansen <i>et al.</i> , 1982	
AR505_0155	Phosphoglycerate kinase	<i>pgk</i>	41.0	Jansen <i>et al.</i> , 1982	
AR505_0474	Phosphoglycerate mutase	<i>apgM</i>	18.0	Jansen <i>et al.</i> , 1982	
AR505_0942	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	<i>gpmA</i>	60.5	Fraser <i>et al.</i> , 1999	
AR505_0470	Phosphopyruvate hydratase	<i>eno</i>	40.3	Jansen <i>et al.</i> , 1982	
AR505_0427,	Phosphoenolpyruvate synthase	<i>ppsA1</i>	38.9	Narindrasorasak & Bridger,	
AR505_1104		<i>ppsA2</i>	18.8	1977	
AR505_1648	Transketolase subunit A	<i>tktA</i>	31.8	Sprenger <i>et al.</i> , 1995	
AR505_1649	Transketolase subunit B	<i>tktB</i>	11.8	Sprenger <i>et al.</i> , 1995	
AR505_1647	Transaldolase	<i>talB</i>	19.5	Sprenger <i>et al.</i> , 1995	
AR505_1685	Ribose-phosphate riphosphokinase	<i>prs</i>	33.4	Hove-Jensen <i>et al.</i> , 1986	
AR505_0461,	Ribulose phosphate epimerase	<i>rpe</i>	40.8	Sprenger, 1995	
AR505_1604		<i>araD</i>	22.2		
AR505_0014	Bifunctional hexulose-6-phosphate synthase/ribonuclease regulator	<i>hxlA</i>	22.2	Yasueda <i>et al.</i> , 1999	
AR505_0177	6-phospho 3-hexuloisomerase	<i>hxlB</i>	30.8	Yasueda <i>et al.</i> , 1999	

\*corresponds to alpha and beta subunit respectively

*TCA cycle.* ISO-H5 is predicted to have an incomplete TCA cycle (Figure 3.7, Table 3.11). ISO4-H5 has the *citZ* gene which encodes citrate synthase required to condense acetyl-CoA and oxaloacetate to produce citrate. Citrate can be further converted to 2-oxoglutarate by aconitate hydratase *AcnA* and isocitrate dehydrogenase *Idh*. In absence of malate dehydrogenase, ISO4-H5 encodes adenylosuccinate lyase *PurB* and adenylosuccinate synthetase *PurA* which convert fumarate to oxaloacetate via aspartate and the broad specificity phosphoserine/aspartate aminotransferases (*SerC/Aat*).

*Gluconeogenesis.* ISO4-H5 encodes all of the genes required for gluconeogenesis from pyruvate to glucose-6-phosphate (Figure 3.7, Table 3.11). Two homologues of the bifunctional phosphoglucose/phosphomannose isomerase (*Pgi1*, *Pgi2*) were found, with *Pgi1* (AR505\_0560) requiring amber codon read-through to translate the full length protein. The ISO4-H5 genome is predicted to harbour two isozymes of phosphoglycerate mutase; 2,3-bisphosphoglycerate-dependent mutase (*GpmA*) and 2,3-bisphosphoglycerate-independent mutase (*ApgM*).

*Pentose phosphate pathway.* The ISO4-H5 genome encodes all the genes required for non-oxidative pentose phosphate pathway and phosphoribosyl pyrophosphate (PRPP) production (Figure 3.7, Table 3.11). The *hxlAB* genes do not exist as an operon in ISO4-H5, and only the *N*-terminal domain of the predicted *HxlA* protein showed homology to characterised proteins.

#### *Amino acid biosynthesis*

The ISO4-H5 is predicted to be able to make 19 amino acids based on the genes present in the genome and the genes involved, their predicted enzymatic activities, and homology to characterised enzymes are summarised in Table 3.12. ISO4-H5 is predicted to be unable to produce methionine and glycine, and as previously mentioned, ISO4-H5 encodes pyrrolysine biosynthesis.

**Table 3.12 Genes encoding amino acid biosynthesis**

Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimentally validated homologue
<b>Glutamate/glutamine</b>				
AR505_0145	Glutamate dehydrogenase	<i>gdhA</i>	55.4	Sakamoto <i>et al.</i> 1975
AR505_0091	Glutamine synthetase	<i>glnA</i>	44.3	Amaya <i>et al.</i> 2005
AR505_0205	Ammonium transporter	<i>atmB</i>	38.2	Soupene <i>et al.</i> 1998
<b>Arginine</b>				
AR505_0674	Ornithine acetyltransferase	<i>argJ</i>	38.2	Hoch, Losick, & Sonenshein, 1993
AR505_0675	Acetylglutamate kinase	<i>argB</i>	27.5	Hoch <i>et al.</i> , 1993
AR505_0673	N-acetylglutamylphosphate reductase	<i>argC</i>	38.0	Hoch <i>et al.</i> , 1993
AR505_0676	Acetylornithine aminotransferase	<i>argD</i>	39.1	Hoch <i>et al.</i> , 1993
AR505_1401	Ornithine carbamoyltransferase	<i>argF</i>	45.4	Hoch <i>et al.</i> , 1993
AR505_0672	Argininosuccinate synthase	<i>argG</i>	51.3	Hoch <i>et al.</i> , 1993
AR505_0671	Argininosuccinase	<i>argH</i>	45.1	Hoch <i>et al.</i> , 1993
AR505_0501, AR505_0504	Carbamoyl-phosphate synthetase	<i>carA</i>	38.6, 40.5	Hoch <i>et al.</i> , 1993
AR505_0500, AR505_0503	Carbamoyl-phosphate synthetase	<i>carB</i>	44.8, 44.2	Hoch <i>et al.</i> , 1993
AR505_0268	Pyruvoyl-dependent arginine decarboxylase	<i>pdaD</i>	29.6	Fukuda <i>et al.</i> 2008
AR505_1607	Agmatinase	<i>speB</i>	31.0	Sekowska <i>et al.</i> 1998
<b>Alanine</b>				
AR505_1199	Alanine aminotransferase	<i>alt</i>	52.3	Ward, Kengen <i>et al.</i> , 2000
AR505_1429	Aspartate/tyrosine/aromatic aminotransferase	<i>aat</i>	36.0	Porat <i>et al.</i> , 2004.
AR505_1666	Aspartate/tyrosine/aromatic aminotransferase	<i>aat</i>	37.3	Porat <i>et al.</i> , 2004.
<b>Aspartate/asparagine</b>				
AR505_0736	Asparagine synthase B1	<i>asnB1</i>	19.2	Scofield <i>et al.</i> , 1990
AR505_1303	Asparagine synthase B2	<i>asnB2</i>	18.7	Scofield <i>et al.</i> , 1990
<b>Lysine</b>				
AR505_0160	Aspartate kinase	<i>lysC</i>	27.7	Yoshioka <i>et al.</i> , 2001
AR505_0491	Aspartate-semialdehyde dehydrogenase	<i>asd</i>	22.9	Paris <i>et al.</i> , 2002
AR505_0161	Dihydrodipicolinate synthase	<i>dapA</i>	29.6	Silk <i>et al.</i> , 1994
AR505_0728	4-hydroxy-tetrahydrodipicolinate reductase	<i>dapB</i>	38.5	Devenish <i>et al.</i> , 2010
AR505_0157	L,L-diaminopimelate aminotransferase	<i>dapL</i>	29.7	Liu <i>et al.</i> , 2010
AR505_0158	Diaminopimelate epimerase	<i>dapF</i>	37.5	Hor <i>et al.</i> , 2010
AR505_0159	Diaminopimelate decarboxylase	<i>lysA</i>	28.7	Andre'O <i>et al.</i> , 2005
<b>Pyrrolysine</b>				
AR505_1325	Pyrrolysine--tRNA ligase	<i>pylS</i>	28.6	Srinivasan <i>et al.</i> , 2002
AR505_1324	Methylornithine synthase	<i>pylB</i>	39.4	Quitterer <i>et al.</i> , 2012
AR505_1323	(2R,3R)-3-methylornithyl- <i>N</i> <sup>6</sup> -lysine synthase	<i>pylC</i>	32.4	Gaston <i>et al.</i> , 2011
AR505_1322	Pyrrolysine synthase	<i>pylD</i>	31.3	Quitterer <i>et al.</i> , 2013
<b>Methionine</b>				
AR505_0293	Homoserine dehydrogenase	<i>metL</i>	22.9	Belfaiza <i>et al.</i> , 1984
AR505_0694	Homoserine <i>O</i> -succinyltransferase	<i>metA</i>	45.4	Born & Blanchard, 1999
AR505_0738	<i>O</i> -acetylhomoserine sulfhydrylase	<i>metI7</i>	43.4, 40.7	Yamagata <i>et al.</i> , 1994, Moore & Thompson, 1967
AR505_0466, AR505_0579	<i>S</i> -adenosylmethionine synthetase	<i>metK1</i> , <i>metK2</i>	14.4, 63.4	Markham <i>et al.</i> , 1980
AR505_0327, AR505_0339, AR505_0340	DNA-cytosine methyltransferase	<i>dcm1</i> , <i>dcm2</i> , <i>dcm3</i>	26.2, 20.5, 18.8	May & Hattman, 1975
AR505_0547, AR505_1788	<i>S</i> -adenosyl-L-homocysteinase	<i>ahcY1</i> , <i>ahcY2</i>	39.9, 39.9	Ogawa <i>et al.</i> , 1987
AR505_1118	<i>S</i> -adenosylhomocystein nucleosidase	<i>mtn</i>	37.6	Cornell & Riscoe, 1998
<b>Threonine</b>				
AR505_0610	Homoserine kinase	<i>thrB</i>	34.7	White, 2003
AR505_0611	Threonine synthase	<i>thrC</i>	54.2	Bult <i>et al.</i> , 1996

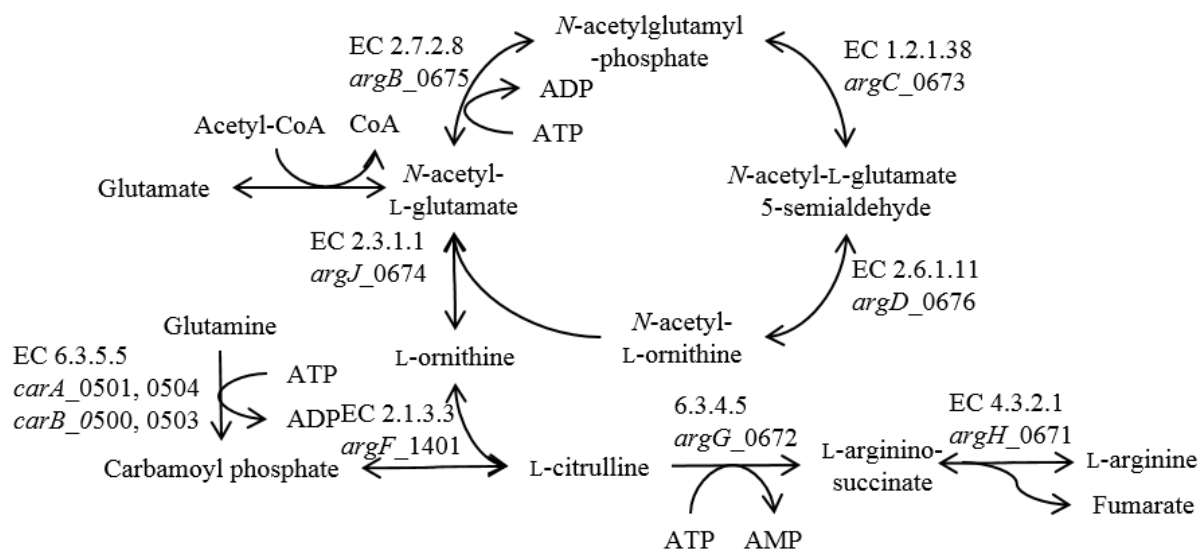
Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimentally validated homologue
<b>Chorismate</b>				
AR505_1148	Triose-phosphate isomerase	<i>tpiA</i>	44.3	White & Xu, 2006
AR505_0508, AR505_0932	Fructose 1,6-bisphosphate aldolase	<i>fba1, fba2</i>	23.8, 21.0	White & Xu, 2006
AR505_0509	Dehydroquinate synthase	<i>aroB</i>	41.8	White, 2004
AR505_0510	Shikimate dehydrogenase	<i>aroE</i>	35.1	Padyana & Burley, 2003
AR505_0511	Shikimate kinase	<i>aroL</i>	14.7	Millar, Lewendon, Hunter, & Coggins, 1986
AR505_0512	3-phosphoshikimate-1-carboxyvinyltransferase	<i>aroA</i>	29.7	Anderson, Sikorski, & Johnson, 1988
AR505_1449	Chorismate synthase	<i>aroC</i>	36.3	White, Millar, & Coggins, 1988
<b>Tryptophan</b>				
AR505_1158	Anthranilate synthase component I	<i>trpE</i>	38.0	Tang <i>et al.</i> , 1999
AR505_1159	Anthranilate synthase component II	<i>trpG</i>	39.7	Tang <i>et al.</i> , 1999
AR505_0997, AR505_1160	Anthranilate phosphoribosyltransferase	<i>trpD1, trpD2</i>	39.2, 40.4	Tang <i>et al.</i> , 1999
AR505_1161	Indole-3-glycerol phosphate synthase	<i>trpC</i>	35.1	Tang <i>et al.</i> , 1999
AR505_1162	Phosphoribosyl anthranilate isomerase	<i>trpF</i>	33.3	Tang <i>et al.</i> , 1999
AR505_1163	Tryptophan synthase beta subunit	<i>trpB</i>	62.3	Tang <i>et al.</i> , 1999
AR505_1164	Tryptophan synthase alpha subunit	<i>trpA</i>	35.4	Tang <i>et al.</i> , 1999
<b>Phenylalanine and tyrosine</b>				
AR505_0490, AR505_1450	Chorismate mutase	<i>aroH1, aroH2</i>	25.8, 20.7	Porat <i>et al.</i> , 2004
AR505_1745	Prephenate dehydrogenase	<i>tyrA</i>	18.1	Porat <i>et al.</i> , 2004
AR505_0515	Prephenate dehydratase	<i>pheA</i>	27.1	Fischer & Jensen, 1987
<b>Proline</b>				
AR505_0108	Glutamate-5-semialdehyde dehydrogenase	<i>proA</i>	56.1	Zhang <i>et al.</i> , 2002
AR505_0109	Glutamate 5-kinase	<i>proB</i>	49.6	J. K. Zhang <i>et al.</i> , 2002
AR505_1634	Pyrroline-5-carboxylate reductase	<i>proC</i>	33.5	Kenklies <i>et al.</i> , 1999
<b>Histidine</b>				
AR505_0494	ATP phosphoribosyl transferase	<i>hisG</i>	32.5	Alifano <i>et al.</i> , 1996
AR505_0486	Phosphoribosyl-ATP pyrophosphatase/phosphoribosyl-AMP cyclohydrolase	<i>hisI</i>	38.9	Alifano <i>et al.</i> , 1996
AR505_0497	[N-(5-phosphoribosyl) formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase	<i>hisA</i>	33.5	Margolies & Goldberger, 1966
AR505_0496	IGP synthase glutamine amidotransferase subunit	<i>hisH</i>	28.4	Demin <i>et al.</i> , 2004
AR505_0487	IGP synthase cycloligase subunit	<i>hisF</i>	41.4	Demin <i>et al.</i> , 2004
AR505_0488	IGP dehydratase	<i>hisB</i>	33.8	Brilli & Fani, 2004
AR505_0495	Histidinol-phosphate aminotransferase	<i>hisC</i>	26.5	Haruyama <i>et al.</i> , 2001
AR505_1770	Histidinol phosphate phosphatase	<i>hisK</i>	13.5	Omi <i>et al.</i> , 2004
AR505_1073	Histidinal dehydrogenase	<i>hisD</i>	40.3	Andorn & Aronovitch, 1982
<b>Serine</b>				
AR505_1664	Phosphoglycerate dehydrogenase	<i>serA</i>	39.2	Ho & Saito, 2001
AR505_0073, AR505_0757	Phosphoserine phosphatase	<i>serB1, serB2</i>	21.7, 12.1	Neuwald & Stauffer, 1985
AR505_1665	Phosphoserine aminotransferase	<i>serC</i>	29.0	Helgadóttir <i>et al.</i> , 2007
<b>Cysteine</b>				
AR505_1192	Serine acetyltransferase	<i>cysE</i>	36.9	Denk & Bock, 1987
AR505_0799	O-acetylserine sulfhydrylase	<i>cysM</i>	40.2	Zhao <i>et al.</i> , 2006
AR505_0695, AR505_0800	Cysteine synthase subunit A	<i>cysK1, cysK2</i>	41.5, 53.6	Flint <i>et al.</i> , 1996



Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimentally validated
<b>Branched chain amino acids</b>				
AR505_1768	Acetolactate synthase small subunit	<i>ilvN</i>	22.1	Xing & Whitman, 1991
AR505_0152, AR505_0153	Acetolactate synthase large subunit	<i>ilvB1</i> , <i>ilvB2</i>	43.1, 42.7	Xing & Whitman, 1991
AR505_0150	2,3-dihydroxy-isovalerate:NADP <sup>+</sup> oxidoreductase	<i>ilvC</i>	46.5	Xing & Whitman, 1991
AR505_1452	Dihydroxy-isovalerate dehydratase	<i>ilvD</i>	56.2	Xing & Whitman, 1991
AR505_1767	Branched-chain amino acid aminotransferase	<i>ilvE</i>	37.2	Xing & Whitman, 1991
AR505_0631	2-isopropylmalate synthase	<i>leuA</i>	36.2	Vartak <i>et al.</i> , 1991
AR505_0531	Isopropylmalate/isocitrate dehydrogenases	<i>idh/leuB1</i>	23.4	Vartak <i>et al.</i> , 1991
AR505_0634	3-isopropylmalate dehydrogenase	<i>leuB2</i>	29.7	Vartak <i>et al.</i> , 1991
AR505_0632	Isopropylmalate dehydratase large subunit	<i>leuC</i>	30.6	Vartak <i>et al.</i> , 1991
AR505_0633	Isopropylmalate dehydratase small subunit	<i>leuD</i>	28.1	Vartak <i>et al.</i> , 1991

*Glutamate/glutamine.* The ISO4-H5 genome is predicted to encode all the genes necessary for glutamate and glutamine biosynthesis using (Table 3.12).

*Arginine.* Arginine biosynthesis via the acetyl cycle pathway, summarized in Figure 3.8 and Table 3.12. The majority of the genes involved in arginine biosynthesis form an operon consisting of *argHGCJBD* (AR505\_0671 – 0676).



**Figure 3.8 Biosynthesis of arginine in ISO4-H5.** Enzymatic reactions are identified by E.C number, gene name followed by locus number. Bifunctional ornithine acetyltransferase/*N*-acetylglutamate synthase (*argJ*). Acetylglutamate kinase (*argB*). *N*-acetylglutamylphosphate reductase (*argC*). Acetylornithine aminotransferase (*argD*). Carbamoyl-phosphate synthetase (*carAB*). Ornithine carbamoyltransferase (*argF*). Argininosuccinate synthase (*argG*). Argininosuccinate lyase (*argH*).

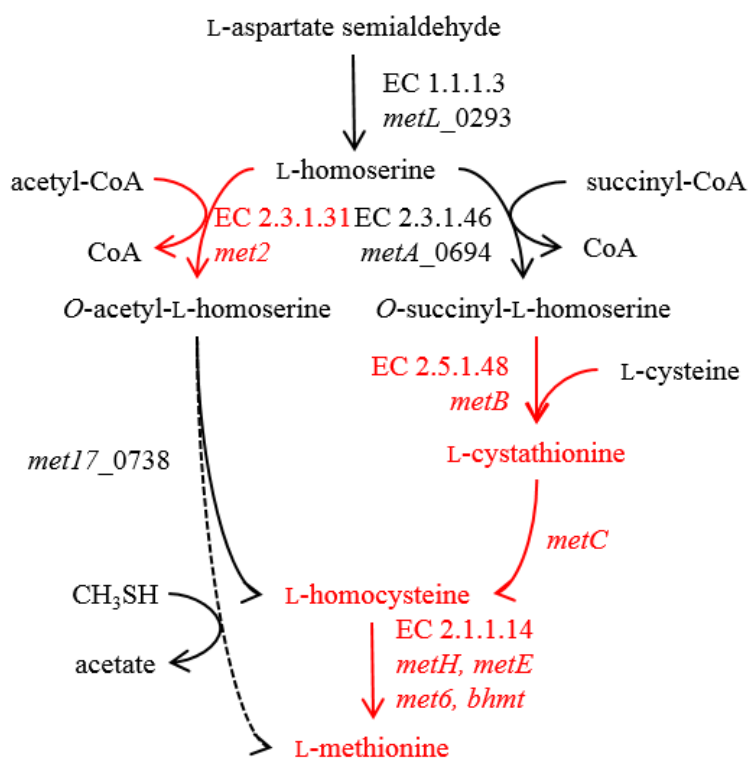
Genes involved in arginine catabolism are also present; arginine decarboxylase (AR505\_0268), which converts L-arginine to agmatine, and agmatinase (AR505\_1607), which breaks down agmatine to putrescine and urea. ISO4-H5 does not encode urea carboxylase gene required to recycle urea.

*Alanine/aspartate/asparagine.* The biosynthesis of alanine involves the reversible transfer of an amino group from glutamate to pyruvate, producing alanine and 2-oxoglutarate (Table 3.12). It is predicted to proceed via the alanine aminotransferase Alt (*alt*, AR505\_1199) a gene specific to alanine, as well as via two broad substrate specificity aspartate aminotransferases, Ast (*aat*, AR505\_1429 and AR505\_1666). The broad substrate specificity aspartate amino transferases also produce aspartate by transamination of oxaloacetate. ISO4-H5 encodes asparagine synthase B, AsnB (*asnB*1, B2; AR505\_0736, 1303) that produces asparagine from aspartate.

*Lysine.* The ISO4-H5 genome is predicted to encode all the genes required for lysine biosynthesis via diaminopimelic acid (DAP) as an intermediate. This pathway branches to homoserine and pyrrolysine biosynthesis downstream.

*Pyrrolysine.* The ISO4-H5 genome is predicted to encode all the genes necessary for pyrrolysine biosynthesis and incorporation.

*Methionine.* Although ISO4-H5 is predicted not to make methionine, the genome encodes a partial methionine biosynthesis pathway, including homoserine dehydrogenase *metL* (AR505\_0293), homoserine *O*-succinyltransferase *metA* (AR505\_0694), and a possible *O*-acetylhomoserine sulfhydrylase (AR505\_0738) gene (Table 3.12, Figure 3.9). Expression of these genes would allow biosynthesis of *O*-succinyl-L-homoserine from L-aspartate semialdehyde, however the ISO4-H5 genome lacks genes encoding the enzymes required to catalyse the remaining steps of methionine biosynthesis. The *metL* in ISO4-H5 only possesses the C-terminal domain of homoserine dehydrogenase (AR505\_0293) and is predicted not to be bifunctional.



**Figure 3.9 Incomplete biosynthesis pathway of methionine in ISO4-H5** Enzymatic reactions are identified by EC number if known, gene name followed by locus number. Pathways and enzymes absent in ISO4-H5 are represented in red text. The dotted line represents a proposed pathway. Homoserine dehydrogenase (*metL*). Homoserine *O*-succinyltransferase (*metA*). Homoserine *O*-acetyltransferase (*met2*). *O*-acetylhomoserine sulfhydrylase (*met17*). *O*-succinylhomoserine(thiol)-lyase (*metB*). Cystathionine  $\beta$ -lyase (*metC*). Cobalamin-independent methionine synthase (*metE*). Methionine synthase (*metH*). Homocysteine methyltransferase (*met6*). Betaine-homocysteine *S*-methyltransferase (*bhmt*).

*Threonine/proline/histidine/serine.* The ISO4-H5 genome is predicted to encode all the genes required for threonine, proline, histidine and serine biosynthesis (Table 3.12).

*Phenylalanine/tyrosine/tryptophan/cysteine.* Homologues of the seven genes required for stepwise synthesis of chorismate (a common precursor in aromatic amino acid biosynthesis) are present in the ISO4-H5 genome (Table 3.12). Based on the genes identified within the genome, ISO4-H5 is predicted to be capable of phenylalanine, tyrosine and tryptophan biosynthesis. The enzymes required for biosynthesis of tryptophan from chorismate are present (Table 3.12). A pseudogene made up of two truncated homologues of *trpB* (AR505\_1151), upstream of the tryptophan biosynthesis operon, is also present.

Homologues of enzymes involved in biosynthesis of phenylalanine and tyrosine from chorismate are scattered throughout the ISO4-H5 genome (Table 3.12). The *pheA* gene in ISO4-H5 carries the N-terminal prephenate dehydratase domain without a C-terminal

regulatory ACT domain, which is named after three proteins containing ACT domain, aspartate kinase, chorismate mutase and TyrA prephenate dehydrogenase.

*Glycine.* Although there are several different ways to produce glycine, (via 5,10-methylenetetrahydrofolate:glycine hydroxymethyltransferase, threonine aldolase, alanine-glyoxylate aminotransferase), no homologues for any of these enzymes were found within the ISO4-H5 genome.

*Isoleucine/leucine/valine.* The pathways of branched chain amino acid biosynthesis in ISO4-H5 appear complete and are similar to what has been found in other organisms, but with minor differences (Table 3.12). ISO4-H5 is unlikely to utilise threonine for branched chain amino acid biosynthesis due to the lack of a homologue to threonine ammonia-lyase. However, the genome contains all the genes necessary for valine biosynthesis from pyruvate (Table 3.12), with 3-methyl-2-oxobutanoate as an intermediate.

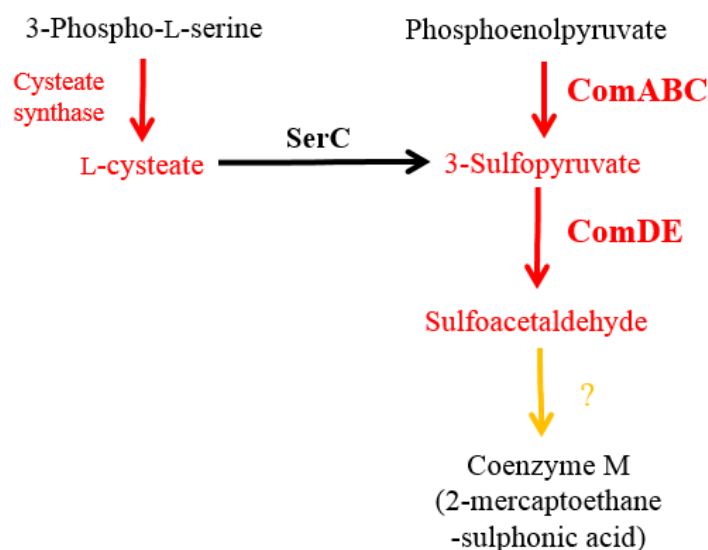
#### *Vitamins and cofactors*

*Cofactor F<sub>430</sub>.* The biosynthesis of tetrapyrrole cofactors such as F<sub>430</sub>, siroheme, and cobalamin all share a common intermediate, uroporphyrinogen. Homologues of genes required for uroporphyrinogen biosynthesis from glutamate are present in the ISO4-H5 genome as a single operon (AR505\_1040 – 1046) in addition to the glutamyl-tRNA synthetase (AR505\_1121) (Table 3.13). The homologue of the uroporphyrinogen-III C-methyltransferase gene *corA* (AR505\_1041) is present in the same operon as the uroporphyrinogen biosynthesis genes in the ISO4-H5 genome. CorA converts uroporphyrinogen-III to precorrin-1, which in turn is converted to precorrin-2 using S-adenyl methionine (SAM) as a cofactor.

**Table 3.13 Genes involved in cofactor biosynthesis**

Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimentally validated homologue
<b>Tetrapyrrole common pathway</b>				
AR505_1040	Uroporphyrinogen III synthase	<i>hemD</i>	18.1	Sasarman <i>et al.</i> , 1987
AR505_1042	Porphobilinogen deaminase	<i>hemC</i>	36.7	Jordan <i>et al.</i> , 1988
AR505_1043	Glutamate-1-semialdehyde-2,1-aminomutase	<i>hemL</i>	44.1	Ilag <i>et al.</i> , 1991
AR505_1044	Delta-aminolevulinic acid dehydratase	<i>hemB</i>	54.0	Bhosale <i>et al.</i> , 1995
AR505_1045	Glutamyl-tRNA reductase	<i>hemA</i>	25.8	Schauer <i>et al.</i> , 2002
AR505_1121	glutamyl-tRNA synthetase	<i>gltX</i>	21.8	Breton <i>et al.</i> , 1986
AR505_1046	Precorrin-2 dehydrogenase	<i>pc2_dh</i>	28.8	Lobo <i>et al.</i> , 2009
<b>F<sub>430</sub></b>				
AR505_1041	Uroporphyrin-III C-methyltransferase	<i>corA</i>	43.3	Blanche <i>et al.</i> , 1991
<b>Riboflavin, FMN, FAD</b>				
AR505_0210	GTP cyclohydrolase	<i>arfA</i>	14.1	Morrison <i>et al.</i> , 2008
AR505_1180	3,4-dihydroxy-2-butanone-4-phosphate synthase	<i>ribB</i>	46.8	Fischer <i>et al.</i> , 2002
AR505_1181	FAD synthetase	<i>ribL</i>	37.6	Mashhadi <i>et al.</i> , 2010
AR505_1182	Riboflavin synthase	<i>ribC</i>	68.4	Fischer <i>et al.</i> , 2004
AR505_1183	6,7-dimethyl-8-ribityllumazine synthase	<i>ribH</i>	55.8	Haase <i>et al.</i> , 2003
AR505_1679	Riboflavin kinase	<i>ribK</i>	38.8	Mashhadi <i>et al.</i> , 2008
<b>NAD<sup>+</sup>/NADP<sup>+</sup></b>				
AR505_0143	L-Aspartate dehydrogenase	<i>aspDH</i>	28.5	Yang <i>et al.</i> , 2003
AR505_1155	Quinolinate synthase	<i>nadA</i>	33.3	Cecilian <i>et al.</i> , 2000
AR505_0471	Quinolinate phosphoribosyltransferase	<i>nadC</i>	16.8	Bhatia & Calvo, 1996
AR505_0377	NH <sub>3</sub> -dependent NAD <sup>+</sup> synthase	<i>nadE</i>	20.9	Ozment <i>et al.</i> , 1999
AR505_0616	Nicotinamide-mononucleotide adenyltransferase	<i>nadM</i>	33.3	Raffaelli <i>et al.</i> , 1997
AR505_0121	NAD kinase	<i>nadK</i>	26.9	Magni <i>et al.</i> , 2006
AR505_1654	Nicotinate phosphoribosyltransferase	<i>pncB</i>	12.0	Wubbolts <i>et al.</i> , 1990

*Cobalamin/thiamine*. The sirohydrochlorin intermediate produced from F<sub>430</sub> biosynthesis can be used for cobalamin (vitamin B12) biosynthesis, (Appendix, Table A.3.1). The ISO4-H5 genome is predicted to encode all but 3 genes required for the anaerobic pathway of cobalamin biosynthesis, the missing genes are *cbiE*, *cbiT* and cobirinic acid-a,c-diamide reductase.



**Figure 3.10 Absence of a complete coenzyme M biosynthesis pathway in ISO4-H5.** Red arrow or names: predicted to be absent in ISO4-H5. Black arrows or names: predicted to be present in ISO4-H5. Yellow arrow or names: unknown enzyme and pathway. Broad specificity phosphoserine aminotransferase (SerC), (R)-

phosphosulfolactate synthase (ComA), phosphosulfolactate phosphatase (ComB), sulfolactate dehydrogenase (ComC), sulfopyruvate decarboxylase (ComDE).

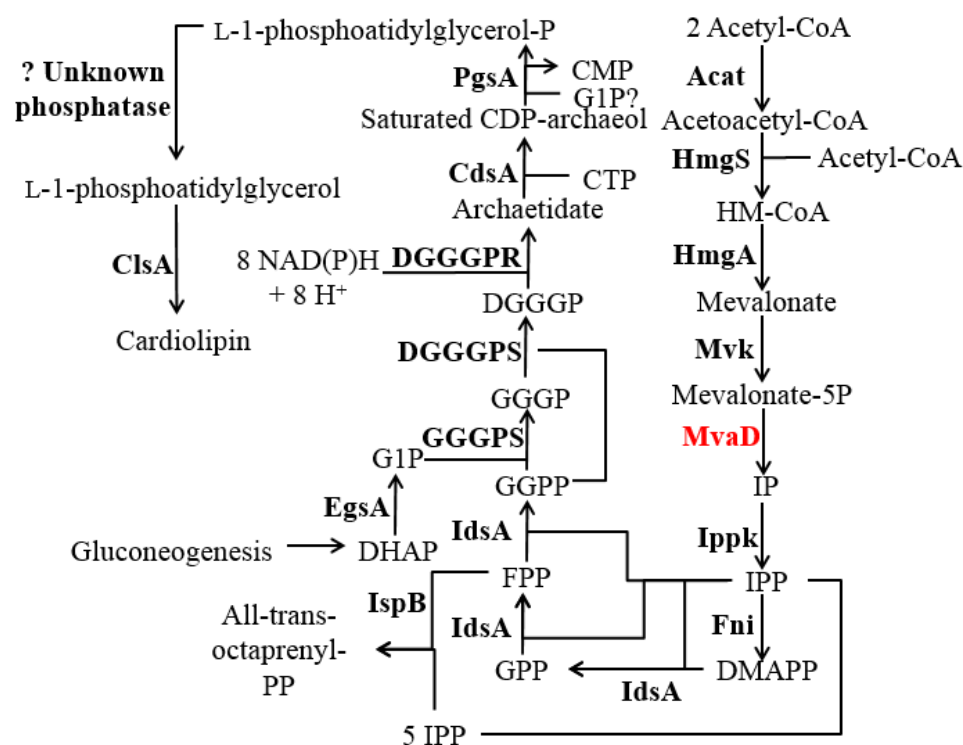
*Riboflavin/FMN/FAD.* Vitamin B<sub>2</sub>, also known as riboflavin, is a precursor to flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The ISO4-H5 genome is predicted to be missing two genes necessary for the biosynthesis of riboflavin; 2-amino-5-formylamino-6-ribosylaminopyrimidin-4(3H)-one 5'-monophosphate deformylase, ArfB, and pyrimidine nucleotide reductase, ArfC, which produces the 5-amino-6-(D-ribitylamino) uracil intermediate (Table 3.13).

*Coenzyme M.* Coenzyme M is an essential cofactor required for methanogenesis, and its biosynthesis from phosphoenolpyruvate requires the (*R*)-phosphosulfolactate synthase gene *comA*, phosphosulfolactate phosphatase gene *comB*, sulfolactate dehydrogenase gene *comC*, sulfopyruvate decarboxylase genes *comDE* and an unidentified enzyme (Figure 3.10). However, homologues to any of the known CoM biosynthesis genes *comABCDE* are absent from the ISO4-H5 genome.

*NAD<sup>+</sup>/NADP<sup>+</sup>.* The biosynthesis of nicotinamide adenine dinucleotide (NAD<sup>+</sup>) and nicotinamide adenine dinucleotide phosphate (NADP<sup>+</sup>) are important cofactors for a large number of enzymes. NAD<sup>+</sup> production in ISO4-H5 is predicted to proceed via two different pathways; from aspartate or from nicotinamide/nicotinic acid (Table 3.13). The ISO4-H5 genome is predicted to encode all but two genes required for the first biosynthesis pathway; aspartate oxidase NadB, that converts aspartate to iminoaspartate, and nicotinate-nucleotide adenylyltransferase NadD, that catalyses the formation of NAD. The proposed NAD biosynthesis pathway via nicotinamide involves NadM, which condenses nicotinamide mononucleotide with ATP to form NAD<sup>+</sup> directly without the requirement of NadE. The nicotinamide mononucleotide is predicted to be produced by a homologue of the nicotinate phosphoribosyltransferase gene (AR505\_1654) from nicotinamide and 5-phospho-D-ribose-1-diphosphate. A homologue of NAD kinase (AR505\_0121) which converts NAD<sup>+</sup> to NADP<sup>+</sup> is also found within the ISO4-H5 genome.

*Biotin.* ISO4-H5 lacks the genes required to synthesise biotin.

In ISO4-H5, glycerol-1-phosphate (G1P) is predicted to be produced from dihydroxyacetone phosphate (DHAP) by EgsA, where the dihydroxyacetone phosphate is derived from the gluconeogenesis pathway by the breakdown of fructose 1,6 phosphate (Figure 3.11). There is one missing step in the conversion of mevalonate-5-phosphate to isopentenyl phosphate due to the lack of phosphomevalonate decarboxylase in the ISO4-H5 genome.



**Figure 3.11 Phospholipid biosynthesis in ISO4-H5.** Enzyme names are displayed in bold. Red enzyme names indicate absence in the ISO4-H5 genome. Acetyl-CoA acetyltransferase (Acat), hydroxymethylglutaryl-CoA synthase (HmgS), hydroxymethylglutaryl-CoA reductase (HmgA), hydroxymethylglutaryl-CoA (HM-CoA), mevalonate kinase (Mvk), phosphomevalonate decarboxylase (MvaD), isopentenyl phosphate (IP), isopentenyl diphosphate (IPP), isopentenyl-diphosphate  $\delta$ -isomerase (Fni), 3,3-dimethylallyldiphosphate (DMAPP), bifunctional short chain isoprenyl diphosphate synthase (IdsA), geranyl diphosphate (GPP), farnesyl diphosphate (FPP), octaprenyl-diphosphate synthase (IspB), geranylgeranyl diphosphate (GGPP), dihydroxyacetone phosphate (DHAP), NAD(P)-dependent glycerol-1-phosphate dehydrogenase (EgsA), glycerol-1-phosphate (G1P), geranylgeranylgeranyl phosphate synthase (GGGPS), geranylgeranylgeranyl phosphate (GGGP), digeranylgeranylgeranyl phosphate synthase (DGGGPS), digeranylgeranylgeranyl phosphate (DGGGP), digeranylgeranylgeranylphospholipid reductase (DGGGPR), CDP-diglyceride synthetase (CdsA), cytidine triphosphate (CTP), cytidine monophosphate (CMP), CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (PgsA), cardiolipin synthase (ClsA).

The proposed *de novo* phospholipid biosynthesis pathway in ISO4-H5 is displayed in Figure 3.11. Based on the presence of the digeranylgeranylgeranylphospholipid reductase (DGGGPR) gene and a CDP-diglyceride synthetase (CdsA) gene, ISO4-H5 is predicted to use archaeetide and saturated CDP-archaeol as intermediates in phospholipid biosynthesis. The

genes in ISO4-H5 suggest that CDP-archaeol proceeds towards cardiolipin biosynthesis, with cardiolipin likely constituting part of the ISO4-H5 cell envelope.

In addition to cardiolipin, ISO4-H5 possesses three geranylgeranyl reductase family proteins, two phosphoglucosamine mutase genes, a cell wall biosynthesis protein Mur ligase family gene, and a NAD-dependent epimerase/dehydratase gene (Table 3.14).

The ISO4-H5 genome does not encode UDP-*N*-acetylglucosamine diphosphorylase, glucosamine-1-phosphate *N*-acetyltransferase and glucosamine-fructose-6-phosphate aminotransferase involved in pseudomureine production. There is a cell wall glycopolymer cluster predicted to be involved in cell envelope biosynthesis (AR505\_0539 - 0561), including 3 genes involved in dTDP-L-rhamnose biosynthesis (AR505\_0552 – 0554). In addition, four genes have been predicted to produce glycerol-manno-heptose (AR505\_1754, AR505\_1770 – 1772).

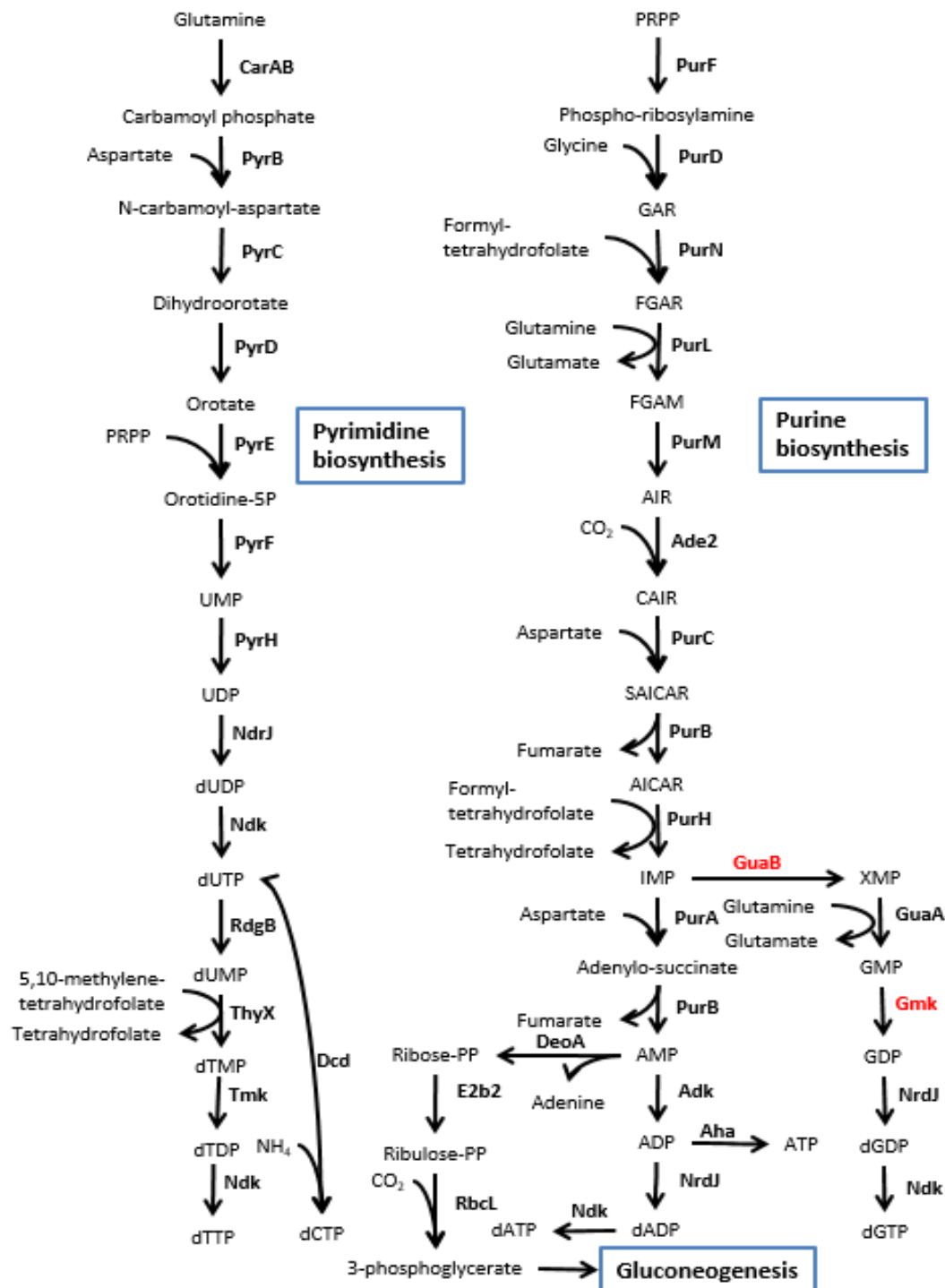


**Table 3.14 Genes involved in lipid biosynthesis**

Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimentally validated homologue
<b>GIP biosynthesis</b>				
AR505_0626	NAD(P)-dependent glycerol-1-phosphate dehydrogenase	<i>egsA</i>	47.4	Nishihara & Koga, 1997
<b>Isoprenoid biosynthesis</b>				
AR505_0004, AR505_0602	Acetyl-CoA acetyltransferase	<i>acat1, acat2</i>	40.4, 19.9	Middleton, 1974
AR505_0601	HMG-CoA synthase	<i>hmgS</i>	23.1	Wilding <i>et al.</i> , 2000
AR505_0768	HMG-CoA reductase	<i>hmgA</i>	47.3	Bochar <i>et al.</i> , 1997
AR505_1431	Mevalonate kinase	<i>mvk</i>	32.8	Primak <i>et al.</i> , 2011
AR505_0192	Isopentenyl phosphate kinase	<i>ippk</i>	31.4	Chen & Poulter, 2010
AR505_0191	Isopentenyl-diphosphate $\delta$ -isomerase	<i>fni</i>	36.4	Yamashita <i>et al.</i> , 2004
AR505_1619	Octaprenyl-diphosphate synthase	<i>ispB</i>	31.3	Asai <i>et al.</i> , 1994
AR505_0190	Bifunctional short chain isoprenyl diphosphate synthase IdsA/GGPS	<i>idsA</i>	39.0	Chen & Poulter, 1994
<b>Phospholipid biosynthesis</b>				
AR505_1588	Geranylgeranyl glyceryl phosphate synthase GGGPS	<i>gggps</i>	43.4	Hemmi <i>et al.</i> , 2004
AR505_1587	Digeranylgeranyl glyceryl phosphate synthase	<i>dgggps</i>	30.8	Zhang <i>et al.</i> , 1990
AR505_1433	Digeranylgeranyl glycerophospholipid reductase	<i>dgggpr</i>	45.4	Nishimura & Eguchi, 2006
AR505_1616	CDP-diglyceride synthetase	<i>cdsA</i>	26.7	Sparrow & Raetz, 1985
AR505_1783	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	<i>pgsA</i>	15.9	Hirabayashi <i>et al.</i> , 1976
AR505_0933	Cardiolipin synthase	<i>clsA</i>	27.7	Ohta <i>et al.</i> , 1985
<b>Cell surface glycopolymer</b>				
AR505_1754	D-sedoheptuloase 7-phosphate isomerase	<i>gmhA</i>	38.8	Kneidinger <i>et al.</i> , 2001
AR505_1771	D-glycero-D-manno-heptose-7-phosphate kinase	<i>hddA</i>	42.4	Kneidinger <i>et al.</i> , 2001
AR505_1770	D-glycero-D-manno-heptose 1,7-bisphosphate 7-phosphatase	<i>gmhB</i>	30.8	Kneidinger <i>et al.</i> , 2001
AR505_1772	D-glycero-D-manno-heptose 1-phosphate guanylyltransferase	<i>hddC</i>	30.8	Kneidinger <i>et al.</i> , 2001
<b>Others</b>				
AR505_0123, AR505_1608	Phosphoglucosamine mutase	<i>glmM1, glmM2</i>	17.7, 30.7	Mengin-Lecreulx & van Heijenoort, 1996
AR505_0892, AR505_1618, AR505_1646	Geranylgeranyl reductase family protein			
AR505_0361	Cell wall biosynthesis protein Mur ligase family			
AR505_0541	NAD-dependent epimerase/dehydratase			

### Nucleic acid metabolism

The biosynthesis of nucleotides requires phosphoribosyl pyrophosphate (PRPP) produced from central carbon metabolism and a formyl group. Other than two genes involved in dGTP biosynthesis, the full complement of genes required for purine and pyrimidine biosynthesis have been predicted in the ISO4-H5 genome (Figure 3.12, Table 3.15, Table 3.16). The IMP dehydrogenase GuaB that produces xanthosine monophosphate (XMP) and guanylate kinase Gmk that phosphorylates GMP to GDP is predicted to be absent.



**Figure 3.12 *De novo* nucleotide biosynthesis in ISO4-H5.** Enzyme names are displayed in bold. Red enzyme names indicate absent in ISO4-H5 genome. Carbamoyl phosphate synthase (CarAB), aspartate transcarbamylase (PyrB), dihydroorotase (PyrC), dihydroorotate dehydrogenase (PyrD), phosphoribose pyrophosphate (PRPP), orotate phosphoribosyltransferase (PyrE), deoxy (d), [adenine (A); cytidine (C); uridine (U); thymidine (T); guanine (G); inosine (I); xanthosine (X)] (Y), [monophosphate (MP); diphosphate (DP); triphosphate (TP)] (ZP) (dYZP), UMP kinase (PyrH), broad substrate specificity nucleoside diphosphate kinase (Ndk), ribonucleoside-diphosphate reductase (NrdJ), dTTP/XTP pyrophosphatase (RdgB), flavin-dependent thymidylate synthase (ThyX), thymidylate kinase (Tmk), CMP/dCMP deaminase (Dcd), amidophosphoribosyl transferase (PurF), phosphoribosylglycine ligase (PurD), (5-phospho-β-D-ribose)glycinamide (GAR), phosphoribosylglycinamide formyltransferase (PurN), formyl-(5-phospho-β-D-ribose)glycinamide (FGAR), phosphoribosylformylglycinamide synthetase (PurL), formamido-5-phospho-β-D-ribose)acetamidine

(FGAM), phosphoribosylformylglycinamide cyclo-ligase (PurM), amino-(5-phospho-ribosyl)imidazole (AIR), phosphoribosylaminoimidazole carboxylase (Ade2), amino-(5-phospho-ribosyl)imidazole-4-carboxylate (CAIR), phosphoribosylaminoimidazole-succinocarboxamide synthase (PurC), 5-phosphoribosyl-4-(*N*-succinocarboxamide)-5-aminoimidazole (SAICAR), bifunctional enzyme adenosuccinate lyase (PurB), amino-1-(5-phosphoribosyl)imidazole-4-carboxamide (AICAR), bifunctional AICAR transformylase/IMP cyclohydrolase (PurH), adenylosuccinate synthetase (PurA), adenylate kinase (Adk), A<sub>1</sub>A<sub>0</sub> ATP synthase (Aha), IMP dehydrogenase (GuaB), GMP synthase (GuaA), guanylate kinase (Gmk), AMP phosphorylase (DeoA), ribulose biphosphate carboxylase (RbcL), ribose-1,5-bisphosphate isomerase (E2b2).

**Table 3.15 Genes involved in pyrimidine biosynthesis and interconversion**

Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimental homologue	validated
<b>Pyrimidine biosynthesis</b>					
AR505_0501, AR505_0504	Carbamoyl phosphate synthase subunit A	<i>carA1</i> , <i>carA2</i>	38.5, 33.4,	Thoden <i>et al.</i> , 1999	
AR505_0500, AR505_0503	Carbamoyl phosphate synthase subunit B	<i>carB1</i> , <i>carB2</i>	48.6, 48.1	Thoden <i>et al.</i> , 1999	
AR505_0437	Aspartate transcarbamylase PyrB	<i>pyrB</i>	48.4	Wang <i>et al.</i> , 2005	
AR505_0436	Aspartate transcarbamylase regulatory subunit	<i>pyrI</i>	31.9	Wang <i>et al.</i> , 2005	
AR505_0281	Dihydroorotase	<i>pyrC</i>	18.7	Thoden <i>et al.</i> , 2001	
AR505_1029	Dihydroorotate dehydrogenase	<i>pyrD</i>	21.5	Norager <i>et al.</i> , 2002	
AR505_1030	Dihydroorotate dehydrogenase electron transfer subunit	<i>pyrK</i>	28.3	Nielsen <i>et al.</i> , 1996	
AR505_1615	Orotate phosphoribosyltransferase	<i>pyrE</i>	26.8	Aghajari <i>et al.</i> , 1994	
AR505_0060	Orotidine-5'-phosphate decarboxylase	<i>pyrF</i>	24.9	Lee & Houk, 1997	
AR505_0251	UMP kinase	<i>pyrH</i>	22.1	Serina <i>et al.</i> , 1995	
AR505_0187	dITP/XTP pyrophosphatase	<i>rdgB</i>	37.6	Chung <i>et al.</i> , 2002	
AR505_1756	Broad substrate specificity nucleoside diphosphate kinase	<i>ndk</i>	45.5	Bernard <i>et al.</i> , 2000	
AR505_0964	Ribonucleoside-diphosphate reductase	<i>nrdJ</i>	25.8	Kolberg <i>et al.</i> , 2004; Torrents <i>et al.</i> , 2000	
AR505_0076	Flavin-dependent thymidylate synthase	<i>thyX</i>	26.6	Myllykallio <i>et al.</i> , 2002	
<b>Interconversion</b>					
AR505_1102, AR505_1690	Thymidylate kinase	<i>tmk1</i> , <i>tmk2</i>	19.0, 31.8	Daws & Fuchs, 1984	
AR505_1827	CMP/dCMP deaminase	<i>dcd</i>	37.3	McIntosh & Haynes, 1986	

ISO4-H5 contains an operon for recycling of AMP, consisting of the genes *deoA*, *rbcL*, and *e2b2* (Table 3.16). The presence of the carbonic anhydrase Cab homologue in the ISO4-H5 genome (AR505\_1245) also adds weight to the functionality of this *RubisCO* operon, as Cab mediates the interconversion of hydrogen carbonate and CO<sub>2</sub>, which usually feeds towards oxaloacetate production.

**Table 3.16 Genes involved in purine biosynthesis and interconversion**

Locus tag	Predicted gene product	Gene	Amino acid identity (%)	Experimental validated homologue
<b>IMP biosynthesis</b>				
AR505_0284	Amidophosphoribosyl transferase	<i>purF</i>	37.6	Messenger & Zalkin, 1979
AR505_0066	Phosphoribosylamineglycine ligase	<i>purD</i>	31.7	Zhang <i>et al.</i> , 2002
AR505_0141	Phosphoribosylglycinamide formyltransferase	<i>purN</i>	17.7	Marolewski <i>et al.</i> , 1994
AR505_1680	Phosphoribosylformylglycinamidine synthetase subunit Q	<i>purQ</i>	37.1	Hoskins <i>et al.</i> , 2004
AR505_1681	Phosphoribosylformylglycinamidine synthetase subunit L	<i>purL</i>	36.5	Hoskins <i>et al.</i> , 2004
AR505_1682	Phosphoribosylformylglycinamidine synthetase subunit S	<i>purS</i>	29.1	Hoskins <i>et al.</i> , 2004
AR505_0037	Phosphoribosylformylglycinamide cyclo-ligase	<i>purM</i>	32.8	Smith & Daum, 1986
AR505_1804	Phosphoribosylaminoimidazole carboxylase	<i>ade2</i>	40.0	Stotz & Linder, 1990
AR505_1595	PurE-related protein		10.9	Meyer <i>et al.</i> , 1992
AR505_0622	Phosphoribosylaminoimidazole-succinocarboxamide synthase	<i>purC</i>	19.0	Meyer <i>et al.</i> , 1992
AR505_0742	Bifunctional enzyme adenosuccinate lyase	<i>purB</i>	25.6	Green <i>et al.</i> , 1996
AR505_1656	Bifunctional AICAR transformylase/IMP cyclohydrolase	<i>purH</i>	50.7	Flannigan <i>et al.</i> , 1990
<b>Adenosine deoxyribonucleotide biosynthesis</b>				
AR505_1168	Adenylosuccinate synthetase	<i>purA</i>	44.1	Honzatko & Fromm, 1999
AR505_0262, AR505_1784	Adenylate kinase	<i>adk1, adk2</i>	44.5, 10.5	Huss & Glaser, 1983
<b>AMP recycling</b>				
AR505_1641	AMP phosphorylase	<i>deoA</i>	47.0	Sato <i>et al.</i> , 2007; Watson <i>et al.</i> , 1999
AR505_1642	Ribulose biphosphate carboxylase RubisCO	<i>rbcL</i>	34.5	Sato <i>et al.</i> , 2007; Watson <i>et al.</i> , 1999
AR505_1643	Ribose-1,5-bisphosphate isomerase	<i>e2b2</i>	42.7	Sato <i>et al.</i> , 2007; Watson <i>et al.</i> , 1999
AR505_1245	Carbonic anhydrase	<i>cab</i>	23.8	Smith & Ferry, 1999
<b>Guanosine deoxyribonucleotide biosynthesis</b>				
AR505_0706, AR505_1802, AR505_1803	GMP synthase	<i>guaA1, guaAa, guaAb</i>	44.8, 11.7, 40.1	Tesmer <i>et al.</i> , 1994

## Cell cycle

ISO4-H5 encodes two origin of replication *orc1/cdc6* genes (AR505\_0001, 1205) and an OB-fold nucleic acid binding domain protein (AR505\_1668) which might act as the single-strand binding protein (SSB) that protects single stranded DNA during replication and recombination. DNA polymerase genes found in the genome include DNA polymerase family D subunits *polD1,2* (AR505\_1438, 1816), family B *polB* (AR505\_0130), and family C *polC* (AR505\_0374). The polymerase is regulated by the sliding clamp, proliferating cell nuclear antigen PcnA (*pcnA*, AR505\_1650) and replication factor C RfcLS genes (*rflC, S*, AR505\_0828, 1209). Genes required for joining of Okazaki fragments are also present, including flap endonuclease *fen* (AR505\_0460), ATP-dependent DNA ligase *dnII* (AR505\_1671) and ribonuclease HII *rnhB* (AR505\_0209). Ribonuclease H (*rnhH*, AR505\_0420) degrades the RNA-DNA hybrid. The tyrosine recombinase XerCD is responsible for termination of replication in bacteria, however no homologous genes are predicted in the ISO4-H5 genome.

The genome segregates in the D phase, and in bacteria the ParAB protein functions to bind the chromosomes and separate them. The ISO4-H5 genome does not harbour ParAB homologs, instead ISO4-H5 relies on chromosome segregation and condensation protein ScpBA (*scpB,A*, AR505\_1609, 1610) and chromosome segregation protein SMC (*smc*, AR505\_1611). There are two additional SMC domain containing proteins (AR505\_0890, 0891) that could also contribute to genome segregation. DNA gyrase GyrAB (*gyrA,B*, AR505\_0033, 0034) and DNA topoisomerase TopAB (*topA,B*, AR505\_0388, 0389, 1299) required for genome segregation has also been found in the genome.

The ISO4-H5 genome is predicted to encode several genes involved in cell division, including two cell division protein genes, *ftsZ* (AR505\_0975, 1792) and the cell division ATPase gene *minD* (AR505\_1412). Cell division is regulated by cell division control proteins Cdc48 (*cdc48-1*, -2, AR505\_0052, AR505\_1810), and two additional genes which might also be related to this function; the AAA family ATPase of Cdc48 subfamily (AR505\_0445) and the cdc-48 related protein (AR505\_0854).

### ***Nucleic acid metabolism***

The ISO4-H5 genome is predicted to encode an archaeal histone (AR505\_1236) and an associated histone acetyltransferase (AR505\_0051) that modifies histones post-translationally. In addition, there are also five GNAT family acetyltransferases and eight SAM-dependent methyltransferase in the ISO4-H5 genome that can possibly modify the histones. The genes predicted to be involved in DNA repair in the ISO4-H5 genome is listed in Table 3.17.

**Table 3.17 DNA recombination and repair**

Locus_tag	Gene product
AR505_0027	8-oxoguanine DNA-glycosylase Ogg
AR505_0028	endonuclease IV
AR505_0117	ssDNA exonuclease RecJ
AR505_0125	replication-associated recombination protein A
AR505_0178	endonuclease IV
AR505_0189	DNA repair and recombination protein RadB
AR505_0329	DNA mismatch endonuclease Vsr
AR505_0391	formamidopyrimidine-DNA glycosylase MutM
AR505_0448	6-O-methylguanine DNA methyltransferase Ogt
AR505_0693	endoribonuclease L-PSP
AR505_0798	ssDNA exonuclease RecJ
AR505_1055	site-specific recombinase
AR505_1056	site-specific recombinase
AR505_1069	DNA repair photolyase
AR505_1084	exodeoxyribonuclease VII small subunit XseB
AR505_1085	exodeoxyribonuclease VII large subunit XseA
AR505_1139	exonuclease
AR505_1154	archaeal Holliday junction resolvase Hjc
AR505_1278	hef nuclease
AR505_1305	endonuclease III Nth
AR505_1309	DNA alkylation repair enzyme
AR505_1402	ssDNA exonuclease RecJ
AR505_1578	excinuclease ABC A subunit UvrA
AR505_1585	DNA repair and recombination protein RadA
AR505_1614	uracil-DNA glycosylase Ung
AR505_1729	resolvase domain-containing protein

### Protein synthesis

**Transcription.** The transcription factors required to initiate and terminate transcription, including the TATA-box binding protein Tbp (*tbp*, AR505\_0740), transcription initiation factor TFIIB (*tfbI*, 2, AR505\_0810, AR505\_1807), transcription factor S, Tfs (*tfs*, AR505\_1651) and NusA-like protein (AR505\_0058). The ISO4-H5 genome encodes 11 subunits of the DNA-directed RNA polymerase (Table 3.18). In addition, there are four RNA-binding proteins (AR505\_0137, 1411, 1425, 1617) that might be involved in transcription regulation.

**Table 3.18 ISO4-H5 RNA polymerase genes**

Locus_tag	Gene product
AR505_0053	DNA-directed RNA polymerase subunit H RpoH
AR505_0054	DNA-directed RNA polymerase subunit B RpoB
AR505_0055	DNA-directed RNA polymerase subunit A' RpoA1
AR505_0056	DNA-directed RNA polymerase subunit A'' RpoA2
AR505_0134	DNA-directed RNA polymerase subunit F RpoF
AR505_0182	DNA-directed RNA polymerase subunit N RpoN
AR505_0922	DNA-directed RNA polymerase subunit K RpoK
AR505_1373	DNA-directed RNA polymerase subunit E'' RpoE2
AR505_1374	DNA-directed RNA polymerase subunit E RpoE
AR505_1439	DNA-directed RNA polymerase subunit D RpoD
AR505_1488	DNA-directed RNA polymerase subunit P RpoP

**Maturation of tRNA.** The tRNA precursor molecules that are transcribed contain 5' and 3' extensions which need to be removed to produce functional tRNAs. The ISO4-H5 genome

encodes genes required this function, including RNase P subunits P29 (*rpp29*, AR505\_0227) and RPR2 (*rpr2*, AR505\_1424), ribonuclease Z (*rnz*, AR505\_0464), tRNA nucleotidyltransferase Cca (AR505\_1583), tRNA intron endonuclease EndA (*endA*, AR505\_1808), 5' RNA ligase (AR505\_1021) and RNA-splicing ligase RtcB (*rtcB*, AR505\_1406). The aminoacyl-tRNA synthetases are responsible for linking amino acids to corresponding tRNAs for amino acids and these are listed in Table 3.19.

**Table 3.19 ISO4-H5 genes involved in aminoacyl-tRNA biosynthesis**

Locus_tag	Gene product
<b>Aminoacyl-tRNA synthetases</b>	
AR505_0198	Alanyl-tRNA synthetase AlaS1
AR505_0200	Aspartate-tRNA synthetase AspS1
AR505_0247	Tryptophanyl-tRNA synthetase TrpS
AR505_0248	Phenylalanyl-tRNA synthetase PheS
AR505_0253	Arginine-tRNA synthetase ArgS
AR505_0387	Lysyl-tRNA synthetase LysS
AR505_0394	Histidyl-tRNA synthetase HisS
AR505_0397	Threonyl-tRNA synthetase ThrS
AR505_0455	Isoleucyl-tRNA synthetase IleS
AR505_0709	Methionyl-tRNA synthetase MetG1
AR505_0710	Methionyl-tRNA synthetase MetG2
AR505_0767	Phenylalanyl-tRNA synthetase PheT
AR505_0972	Leucyl-tRNA synthetase LeuS
AR505_1121	Glutamyl-tRNA synthetase GltX
AR505_1150	Aspartate-tRNA synthetase AspS2
AR505_1191	Cysteiny-tRNA synthetase CysS
AR505_1325	Pyrrolysine-tRNA synthetase PylS
AR505_1339	Valyl-tRNA synthetase ValS
AR505_1341	Glycyl-tRNA synthetase GlyS
AR505_1426	Seryl-tRNA synthetase SerS
AR505_1448	Alanyl-tRNA synthetase AlaS2
AR505_1494	Tyrosyl-tRNA synthetase TyrS
AR505_1684	Prolyl-tRNA synthetase ProS
<b>ATP dependent transamination</b>	
AR505_0068	Glutamyl-tRNA(gln) amidotransferase GatD
AR505_0069	Glutamyl-tRNA(gln) amidotransferase GatE
AR505_0201	Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase GatB
AR505_0202	Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase GatA
AR505_0203	Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase GatC

**Maturation of rRNA.** The endonuclease and ribonuclease involved in tRNA maturation are also involved in rRNA maturation. In addition, the ISO4-H5 genome encodes a ribosome maturation protein SBDS (AR505\_1483) and four genes encoding exosomes subunits (AR505\_1358), RNA-binding protein Rrp4 (*rrp4*, AR505\_1484), exonuclease Rrp41 (*rrp41*, AR505\_1485) and RNA-binding protein Rrp42 (*rrp42*, AR505\_1486).

**RNA modifications.** Genes involved in RNA post-transcriptional modifications, including tRNA modification to wyosine, queuosine and archaeosine are shown in Table 3.20. There are no functional snRNA nor RNA components of C/D box methylation guide ribonucleoprotein complex identified within the ISO4-H5 genome, although a aNOP56 subunit (AR505\_1605) and fibrillarin (AR505\_0120) are found.

**Table 3.20 ISO4-H5 genes involved in aminoacyl-tRNA biosynthesis**

Locus_tag	Gene product
<b>Wyosine, queuosine and archaeosine</b>	
AR505_0438	Archaeosin tRNA-ribosyltransferase TgtA1
AR505_0637	Archaeosin tRNA-ribosyltransferase TgtA2
AR505_0827	Wyosine biosynthesis protein TYW1
AR505_1414	7-cyano-7-deazaguanosine biosynthesis protein QueE
AR505_1415	Queuosine biosynthesis protein QueC
AR505_1416	6-pyruvoyl tetrahydropterin synthase QueD
<b>Others</b>	
AR505_0038	Ribosomal RNA large subunit methyltransferase J RrmJ
AR505_0044	tRNA 2'-O-methylase (predicted pseudogene)
AR505_0065	tRNA pseudouridine synthase A TruA
AR505_0132	Dimethyladenosine transferase KsgA
AR505_0243	tRNA pseudouridine synthase B TruB
AR505_0307	SAM-dependent methyltransferase
AR505_0536	SAM-dependent methyltransferase
AR505_0809	tRNA-dihydrouridine synthase DusA
AR505_0973	SAM-dependent methyltransferase
AR505_0981	SAM-dependent methyltransferase
AR505_1008	SAM-dependent methyltransferase
AR505_1146	Ribosomal-protein-alanine acetyltransferase RimI
AR505_1196	SAM-dependent methyltransferase
AR505_1203	SAM-dependent methyltransferase
AR505_1233	Pseudouridylate synthase
AR505_1275	tRNA pseudouridine synthase D TruD
AR505_1285	SAM-dependent methyltransferase
AR505_1460	MiaB-like tRNA modifying enzyme
AR505_1586	N <sub>2</sub> , N <sub>2</sub> -dimethylguanosine tRNA methyltransferase Trm
AR505_1658	Ribosomal protein L11 methyltransferase PrmA
AR505_1683	tRNA methyltransferase subunit
AR505_1791	RNA methyltransferase TrmH family

*Ribosomal proteins.* The gene encoding proteins in the large (50S) and small (30S) subunits of the ribosome are listed in Table A.3.1.

*Translation factors.* Translation factors are essential in facilitation of translational initiation, peptide elongation and termination and those found in ISO4-H5 are listed in Table 3.21.

**Table 3.21 Translation factors in the ISO4-H5 genome**

Locus_tag	Gene product
<b>Translation initiation factors</b>	
AR505_0139	Translation initiation factor aIF-1A
AR505_0226	Translation initiation factor aSUI1
AR505_0782	Translation initiation factor aIF-6
AR505_1459	Translation initiation factor aIF-2 alpha subunit
AR505_1516	Translation initiation factor aIF-2 alpha subunit
AR505_1603	Translation initiation factor aIF-2 beta subunit
AR505_1606	Translation initiation factor aIF-5A
AR505_1752	Translation initiation factor aIF-2 gamma subunit
AR505_1755	Translation initiation factor aIF-2
<b>Translation elongation factors</b>	
AR505_1011	Elongation factor Tu domain
AR505_1454	Translation elongation factor aEF-1 alpha subunit
AR505_1455	Translation elongation factor aEF-2
AR505_1774	Translation elongation factor aEF-1 beta
<b>Translation termination factors</b>	
AR505_0252	Peptide chain release factor aRF1



*Degradation of mRNA.* In addition to the aforementioned exosome complex that facilitates degradation of defective mRNAs, a NMD3 family protein (AR505\_0020) involved in nonsense-mediated mRNA decay is predicted in the ISO4-H5 genome.

### ***Protein fate***

*Secretion.* The components of ISO4-H5 secretion systems are listed in Table 3.22.

**Table 3.22 Secretion in the ISO4-H5 genome**

<b>Locus_tag</b>	<b>Gene product</b>
AR505_0006	Signal recognition particle receptor FtsY
AR505_0758	Signal recognition particle protein SRP54
AR505_1473	Signal recognition particle protein SRP19
AR505_0386	Preprotein translocase subunit SecG
AR505_0241	Preprotein translocase subunit SecY
AR505_1799	Preprotein translocase subunit SecE
AR505_0667, AR505_1612, AR505_1828	Signal sequence peptidase SPI

*Protein folding.* A variety of chaperones are encoded by ISO4-H5 (Table 3.23). A nascent polypeptide-associated complex protein (alpha-NAC homolog, AR505\_0195) involved in stabilizing newly synthesised polypeptides is predicted. Two subunits of prefoldin PfdAB (*pfdAB*, AR505\_0007, 1488) are predicted and are thought to act to transfer non-native polypeptides to chaperones. The molecular chaperones then hold the polypeptides in place and assist it to fold into its native state. Another class of enzymes that assist protein folding are the peptidyl-prolyl isomerases (AR505\_0673, AR505\_1028, AR505\_1235). A class of chaperonin specific to archaea, the thermosome, was also seen. This high molecular weight complex is formed from only two subunits (AR505\_0092, AR505\_1062), which binds denatured polypeptides and functions as a chaperonin.

**Table 3.23 ISO4-H5 genes involved in protein folding and stability**

<b>Locus_tag</b>	<b>Gene product</b>
<b>Prefoldin</b>	
AR505_0007	Prefoldin alpha subunit PfdA
AR505_1488	Prefoldin beta subunit PfdB
<b>Chaperones</b>	
AR505_0093	Molecular chaperon GrpE
AR505_0094	Chaperon protein DnaK1
AR505_0095	Chaperon protein DnaJ
AR505_0286	Heat shock protein Hsp90
AR505_1228	Chaperon protein DnaK2
AR505_1232	Chaperon protein DnaK3
AR505_1248	ATP-dependent chaperon protein ClpB
<b>Peptidyl-prolyl isomerase</b>	
AR505_0685	Peptidyl-prolyl cis-trans isomerase FKBP-type
AR505_1028	Peptidyl-prolyl cis-trans isomerase
AR505_1235	Peptidyl-prolyl cis-trans isomerase
<b>Thermosome</b>	
AR505_0092	Thermosome subunit
AR505_1062	Thermosome subunit

*Protein degradation.* Misfolded, denatured proteins and spent regulatory proteins need to be degraded, as they could cause aggregation within the cell. The degradation of proteins in ISO4-H5 is mediated by an array of peptidases and a proteasome (Table 3.24)

**Table 3.24 ISO4-H5 genes involved in protein degradation**

Locus_tag	Gene product
<b>Proteasome</b>	
AR505_0967	Proteasome endopeptidase complex
AR505_1482	Proteasome alpha subunit PsmA
AR505_1674	Proteasome-activating nucleotidase
<b>Peptidase (transmembrane)</b>	
AR505_0013	Peptidase M48 family
AR505_0264	Peptidase M50 family
AR505_0615	ATP-dependent protease S16 family
AR505_0702	Peptidase C1A papain
AR505_0721	CAAX amino terminao protease family protein
AR505_0785	Peptidase M50 family
AR505_1813	CAAX amino terminao protease family protein
<b>Peptidase (cytoplasmic)</b>	
AR505_0122	Metal-dependent protease
AR505_0144	Peptidase M24 family
AR505_0270	Peptidase M16 family
AR505_0449	Aminoacyl-histidine dipeptidase PepD
AR505_0907	Peptidase U62, modulator of DNA gyrase
AR505_0908	Peptidase U62, modulator of DNA gyrase
AR505_0919	Proline-specific peptidase
AR505_1001	Peptidase U32 family
AR505_1151	ATP-dependent protease
AR505_1318	Peptidase M18 family
AR505_1443	Methionine aminopeptidase Map

## *Regulation*

The ISO4-H5 genome is predicted to encode 37 predicted transcriptional regulators, most with unknown functions (Table A.3.1).

**Table 3.25 ISO4-H5 genes involved regulatory protein interaction**

<b>Locus tag</b>	<b>Gene product</b>
AR505_0138	Serine/threonine protein kinase RIO1 family
AR505_0206	Nitrogen regulatory protein P-II
AR505_0259	Phosphate uptake regulator PhoU1
AR505_0260	Phosphate uptake regulator PhoU2
AR505_0402	Nitrogen regulatory protein P-II
AR505_0609	Low molecular weight phosphotyrosine protein phosphatase
AR505_1176	Regulatory protein LytS
<b>TPR repeat containing proteins</b>	
AR505_0204	TPR repeat-containing protein
AR505_0275	TPR repeat-containing protein
AR505_0366	TPR repeat-containing protein
AR505_0367	TPR repeat-containing protein
AR505_0368	TPR repeat-containing protein
AR505_0370	TPR repeat-containing protein
AR505_0373	TPR repeat-containing protein
AR505_0419	TPR repeat-containing protein
AR505_0536	TPR repeat-containing protein
AR505_0557	TPR repeat-containing protein
AR505_0641	TPR repeat-containing protein
AR505_0793	TPR repeat-containing protein
AR505_0795	TPR repeat-containing protein
AR505_0898	TPR repeat-containing protein
AR505_0899	TPR repeat-containing protein
AR505_0917	TPR repeat-containing protein
AR505_1020	TPR repeat-containing protein
AR505_1048	TPR repeat-containing protein
AR505_1145	TPR repeat-containing protein
AR505_1212	TPR repeat-containing protein
AR505_1213	TPR repeat-containing protein
AR505_1214	TPR repeat-containing protein
AR505_1216	TPR repeat-containing protein
AR505_1225	TPR repeat-containing protein
AR505_1273	TPR repeat-containing protein
AR505_1334	TPR repeat-containing protein
AR505_1351	TPR repeat-containing protein
AR505_1463	TPR repeat-containing protein

## *Nitrogen metabolism*

ISO4-H5 is capable of utilising nitrogen via ammonium uptake and glutamate/glutamine biosynthesis, as mentioned previously, but also has an array of genes associated with nitrogen fixation (Table 3.26). However the involvement of these genes in nitrogen fixation activity is not clear.

**Table 3.26 ISO4-H5 genes potentially involved nitrogen fixation**

Locus tag	Gene product
AR505_0186	Nitroreductase family protein
AR505_0207	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase
AR505_0359	Nitrogenase iron protein NifH1
AR505_0591	Nitroreductase family protein
AR505_0822	Nitroreductase family protein
AR505_0956	Hydroxylamine reductase Hcp
AR505_1059	FeS cluster assembly scaffold protein NifU
AR505_1060	Cysteine desulfurase NifS
AR505_1268	Dinitrogenase iron-molybdenum cofactor biosynthesis protein
AR505_1288	Oxidoreductase/nitrogenase component 1
AR505_1289	Nitrogenase
AR505_1413	Dinitrogenase iron-molybdenum cofactor biosynthesis protein
AR505_1447	Nitrogenase cofactor biosynthesis protein NifB
AR505_1501	Nitrogenase iron protein NifH2
AR505_1502	Oxidoreductase/nitrogenase component 1

### ***Oxidative stress***

Under anaerobic conditions, oxygen is not the terminal electron acceptor, so the formation of reactive oxygen species is not problematic, and explains the lack of genes such as catalases and superoxide dismutases (SOD) in ISO4-H5. However, rumen methanogens require genes to deal with oxidative stress (Table 3.27). ISO4-H5 encodes a desulfoferrodoxin (*dfx*, AR505\_0630) which detoxifies oxygen to hydrogen peroxide, and a FAD linked oxidase domain-containing protein (AR505\_0263), which may detoxify oxygen. The hydrogen peroxide is detoxified by rubrerythrin (*rbr*, AR505\_1353) to water via NADH peroxidase activity. Like most archaea, the ISO4-H5 genome lacks genes encoding the glutathione pathway, but it harbors the thiol redox pathway of thioredoxin (*trxA*, AR505\_0047) and thioredoxin-disulfide reductase TrxB (*trxB*, AR505\_0048).

**Table 3.27 ISO4-H5 genes involved in oxidative stress**

Locus tag	Gene product
AR505_0212	CoA-disulfide reductase cdr
AR505_0263	FAD linked oxidase domain-containing protein
AR505_0630	Desulfoferrodoxin Dfx
AR505_0700	Rubredoxin
AR505_0766	Peroxiredoxin AhpC
AR505_0970	Oxidoreductase
AR505_1047	Rubredoxin Rbr
AR505_1353	Rubrerythrin Rbr
<b>Thioredoxin</b>	
AR505_0047	Thioredoxin TrxA
AR505_0048	Thioredoxin-disulfide reductase TrxB1

### 3.3 Discussion

This chapter describes the sequencing, gap closing and bioinformatics analysis of the ISO4-H5 genome. ISO4-H5 is the first reported member of the Methanomassiliicoccales isolated from the ovine rumen, and its genome sequence represents an important contribution to the knowledge accumulating about this newly discovered group of rumen methanogens. Its 1.9 Mb genome and the analysis of its gene expression under different growth conditions has provided new insight into the lifestyle and cellular processes of this rumen methanogen.

A reported feature of genomes from members of the Methanomassiliicoccales order is the lack of organisation of rRNA genes into an operon (Borrel *et al.* 2014). rRNA genes are typically ordered 16S, 23S and 5S and are organised in an operon structure (Brosius *et al.* 1981). Archaeal rRNA operons vary in both structure and frequency (Dennis 1997), with 50% of sequenced archaeal genomes containing only one rRNA operon (Yip *et al.* 2013). However, the number of copies of rRNA operons within archaeal genomes can be as high as four (Yip *et al.* 2013), with some rRNA operons containing one or two tRNAs within spacer sequences, including tRNA<sup>Glu</sup>, tRNA<sup>Ile</sup>, tRNA<sup>Ala</sup> and tRNA<sup>Cys</sup> (Lund *et al.* 1976; Ikemura and Nomura 1977; Morgan *et al.* 1977; Keller *et al.* 1980; Chant and Dennis 1986). Some archaeal genomes are found with the 5S rRNA gene existing outside the rRNA operon, as observed in the *Desulfurococcus mobilis*, *Sulfolobus solfataricus* and *Methanococcus jannaschii* genomes (Kjems and Garrett 1987; Bult *et al.* 1996; She *et al.* 2001). Rumen methanogens typically have their rRNA genes organised in an operon, however, the ISO4-H5 genome lacks an rRNA operon. This is similar to the *Thermoplasma acidophilum* genome (Tu and Zillig 1982) from the order Thermoplasmatales, to which members of the Methanomassiliicoccales were originally assigned (Tajima *et al.* 2001). While the separation of the 16S, 23S and 5S rRNA genes appears to be a trait shared by all members of the Methanomassiliicoccales (Borrel *et al.* 2014) the reason for this is not clear. The cellular rRNA content determines the rate of protein synthesis and cytoplasmic mass accumulation, the location of rRNA operon is on average 10 minutes from the origin of replication in an *E. coli* with a doubling time of 100 minutes (Bremer and Dennis 2008). It is conceivable that the closer the rRNA genes to the origin of replication, the earlier it would produce rRNA and facilitate protein synthesis, therefore the location of rRNA genes may be correlated with the growth rate of the organism (Bremer and Dennis 2008). There are seven rRNA genes in ISO4-H5, and their average distance to the origin of replication is 16% of the genome size, the longer distance may suggest the ISO4-H5 could possibly be a

slow growing organism. The genome of ISO4-H5 encodes 47 tRNAs encoding for all amino acids. Introns in transfer RNAs are found in all three kingdoms of life and present on average in 15% of archaeal tRNAs (Yoshihisa 2014). Three of the ISO4-H5 tRNAs; tRNA<sup>Met</sup>, tRNA<sup>Trp</sup> and tRNA<sup>Tyr</sup>, contain an intron. The tRNA intron endonuclease *endA* (AR505\_1808) required for tRNA intron removal (Thompson *et al.* 1989) has been identified within the ISO4-H5 genome. A difference observed in the tRNA<sup>Met</sup> intron in ISO4-H5 is that it does not encode a C/D box snRNA, which is thought to be involved in *O*-methylation of rRNA in *Nanoarchaeum equitans* (Randau *et al.* 2005). This suggests that either *O*-methylation of rRNA does not occur in ISO4-H5 or it might be carried out by other means. The ISO4-H5 genome is predicted to encode only single copies of tRNA<sup>Trp</sup> and tRNA<sup>Tyr</sup>, therefore, intron removal is essential to allow the synthesis of their mature tRNAs needed for translation. The pseudouridine modification of tRNA<sup>Tyr</sup> is dependent on its intron in *Saccharomyces cerevisiae* (Johnson and Abelson 1983), therefore the intron in tRNA<sup>Tyr</sup> of ISO4-H5 may perform a similar role, because ISO4-H5 encodes genes required for pseudouridylation within its genome. The tRNA<sup>Trp</sup> intron has been observed to perform a catalytic function in *Pyrococcus abyssi*, where it acts as a snoRNA and guides the ribose methylation of tryptophan pre-tRNA exons (Clouet d'Orval *et al.* 2001). However, the ISO4-H5 tRNA<sup>Trp</sup> intron sequence does not match any known sequence in the Rfam database (Griffiths-Jones *et al.* 2005) and whether the ISO4-H5 intron in tRNA<sup>Trp</sup> harbors catalytic activity remains unknown.

Horizontal gene transfer occurs commonly amongst prokaryotes, and continuous acquisition and removal of genes in the genome is an integral part of genome evolution (Ochman *et al.* 2000; Koonin *et al.* 2001), assisting with environmental adaptation (Gilbert and Cordaux 2013). The genes acquired from horizontal transfer often differ to other genes in the genome in base composition, codon usage and bias of di- and tri-nucleotides (Ochman *et al.* 2000). Two regions in the ISO4-H5 genome are considered horizontally acquired based on their atypical base composition (Table 3.09). The first region (Region 7) contains genes involved in dTDP-L-rhamnose biosynthesis (AR505\_0552 – 0554). In Gram-negative bacteria, rhamnose is a component of the outer membrane exopolysaccharide (EPS) as well as the large outer membrane lipopolysaccharide known as the *O*-antigen (Samuel and Reeves 2003; Jofre *et al.* 2004). Rhamnose has been detected in EPS in the methanogens *Methanobacterium formicum* and *Methanosarcina mazei* (Veiga *et al.* 1997). This suggests that rhamnose may play a role in the cell envelope structure of Methanomassiliicoccales. A second region (Region 4) contains a large cluster of hypothetical genes, but a closer examination identifies a 25 Kb section with

25 predicted ORFs (AR505\_0333 to AR505\_0358) which show some characteristics of a prophage (Canchaya *et al.* 2003). Region 4 also contains a phage integrase (AR505\_0313) which is required for phage integration, and three DNA-cytosine methyltransferases genes (AR505\_0327, AR505\_0339, AR505\_0340) required to circumvent host restriction-modification barriers and increase likelihood of successful infection (Yasmin *et al.* 2010); (Buhk *et al.* 1984). However, no phage structural components were discernable within this region, so it is unclear whether Region 4 represents an integrated prophage, phage remnant or other mobile element.

Insertion sequence (IS) elements and other mobile elements contribute to plasticity and evolution of archaeal genomes, and are capable of facilitating horizontal transfer (Brügger *et al.* 2002). IS elements are segments of DNA, less than 2.5 Kb in length, and capable of inserting into a target sequence (Frost *et al.* 2005). IS elements often encode a transposase flanked with short inverted repeats necessary for transposase recognition and cleavage (Mahillon and Chandler 1998). A region of DNA flanked by two IS elements may be mobilized as a composite transposon (Siguier *et al.* 2015). IS elements constitute 2.3% of the ISO4-H5 genome, whereas IS elements were found to constitute only 1.2% of the Mx1201 genome (Borrel *et al.* 2012). Higher number of IS elements may act as a driving force of genome evolution (Biemont and Vieira 2006), aiding ISO4-H5 to persist and define its niche within the competitive habitat of the rumen environment.

The ISO4-H5 genome is predicted to encode two toxin and four antitoxin system genes (Table A.3.1). Toxin/antitoxin systems are typically identified within plasmids, where they trigger the death of daughter cells that fail to inherit the plasmids (Hayes 2003). Toxin/antitoxin genes have also been identified within chromosomes where they function in programmed cell death and in metabolic control during nutritional stress (Gerdes *et al.* 2005; Allocati *et al.* 2015). The ISO4-H5 genome has one example of toxin and antitoxin genes being co-located which suggests they function together as a system (Gerdes *et al.* 2005). The predicted toxin gene (AR505\_0858) is 34.4% homologous to the predicted toxin gene *relE* (WP\_011992585.1) from the lambdaSa04 prophage found in *Campylobacter curvus* 525.92 (Fouts *et al.* 2007). Similarly, the antitoxin gene is 29.8% homologous to the prevent-host-death protein (WP\_009650570.1) from the same prophage and organism. The function of the RelE protein in bacteria and archaea is to cleave the mRNA at the ribosome and trigger non-lethal arrest of translation that inhibits cell growth, and its function can be triggered by nutrient starvation (Christensen and Gerdes 2003). There is no genomic evidence to suggest presence of a plasmid

in the ISO4-H5, which suggests the toxin/antitoxin may function as a regulator to halt cellular growth. There is a TetR family regulator with a predicted transmembrane helice immediately downstream of the *relE* gene. The TetR family regulator HapR in *Vibrio cholera* regulates virulence genes, HapR has been found to be regulated by quorum-sensing signals (Miller *et al.* 2002). Quorum sensing may converge with starvation-sensing (Lazazzera 2000). The TetR family regulator gene downstream of *relE* gene may function similary, regulated by quorum/starvation-sensnig signals and regulating the toxin/antitoxin system.

The ISO4-H5 genome is also predicted to contain a CRISPR element. CRISPR elements typically confer acquired immunity against phage and conjugative plasmids (Marraffini and Sontheimer 2010). The CRISPR elements function in a manner similar to RNA interference; the spacer sequence between repeat elements contains DNA sequence from former invading genetic elements. Spacer regions can uptake new material to match new foreign genetic elements and prevent conjugation and transduction of foreign genetic elements (Marraffini and Sontheimer 2008). The spacer sequences of the identified CRISPR element were used to identify foreign genetic elements that ISO4-H5 may have encountered in the past (Biswas *et al.* 2013; Shariat *et al.* 2015). Out of 32 spacer sequences analysed, 17 matched known phage and plasmid sequences. Plasmid hosts included *Azospirillum* spp. and *Rahnella* spp., both genera are nitrogen-fixing bacteria found in the rhizosphere of grass, soybeans or wheat (Berge *et al.* 1991; Steenhoudt and Vanderleyden 2000; Weidner *et al.* 2012). It is conceivable that within the rumen environment ISO4-H5 may have come into contact with similar organisms colonising forage plant material, and may have been exposed to their mobile genetic elements.

The defining metabolism of the Methanomassiliicoccales is their formation of methane, and like others members of this order (Borrel *et al.* 2014), ISO4-H5 possesses a complete set of genes required to carry out methylotrophic methanogenesis using methyl substrates. The methyl groups are transferred onto CoM by methyltransferases and then reduced to methane by the Mrt complex (Borrel *et al.* 2014). This particular pathway of methylotrophic methanogenesis appears to be unique to the Methanomassiliicoccales order (Borrel *et al.* 2014). Other methylotrophic methanogens within the rumen, such as *Methanosarcina* spp. disproportionate methanol by electron bifurcation, oxidizing it to produce CO<sub>2</sub> while generating reducing potential to reduce the methyl group to methane (Welander and Metcalf 2005). The methanogenesis pathway in ISO4-H5 lacks the genes encoding the enzymes required to oxidize methanol to CO<sub>2</sub>, and is predicted to only reduce methylated compounds directly to methane. Functionally, this is similar to *Methanosphaera stadtmanae* MCB-3,



which encodes all the genes for the enzymes needed to oxidize methanol to CO<sub>2</sub> but does not use this pathway due to the lack of genes encoding synthesis of molybdopterin, a cofactor required for formation of an active formylmethanofuran dehydrogenase (Fricke *et al.* 2006). The Methanomassiliicoccales is the second most abundant order of rumen methanogens (Janssen and Kirs 2008), which suggests that ISO4-H5 and other rumen Methanomassiliicoccales compete successfully against *Methanosarcina* spp. and *Methanosphaera* spp. for methyl substrates. ISO4-H5 strict depend on methylotrophic methanogenesis and direct reduction of methyl-substrates, theoretically ISO4-H5 can produce one mole of CH<sub>4</sub> from one mole of H<sub>2</sub>, whereas hydrogenotrophic methanogens require four moles of H<sub>2</sub> to produce one mole of CH<sub>4</sub>, this suggests ISO4-H5 likely only requires a small amount of hydrogen to survive, which could provide the rumen Methanomassiliicoccales an unique advantage and allowed establishment of their niche within the rumen environment. While *Methanosphaera* spp. also utilises methylotrophic methanogenesis, they lack the specialization of ISO4-H5 as they carry the full array of hydrogenotrophic methanogenesis genes, which implies additional cellular expenditure, this may explain why members of Methanomassiliicoccales order are more abundant than *Methanosphaera* spp. in the rumen.

Based on the genome sequence, ISO4-H5 is predicted to utilise methanol, mono-, di-, tri-methylamine and methylthiol as substrates for methanogenesis. The ability to utilise five different substrates for methanogenesis offers ISO4-H5 a competitive advantage over the methylotrophic rumen methanogen *Methanosphaera* which is limited to the use of methanol as a methyl substrate (Jeyanathan, 2010). Three methyltransferase genes are typically required to transfer the methyl group from the substrate onto CoM, and to utilise methanol ISO4-H5 contains the genes, *mtaA*, *mtaB*, and *mtaC*. MtaB is a methanol-specific corrinoid methyltransferase which transfers the methyl group from methanol onto the corresponding corrinoid protein component, MtaC. MtaA is a methylcobalamin:coenzyme M methyltransferase responsible for transferring the methyl group from the corrinoid protein onto the CoM for methanogenesis. Similar sets of genes are identified within the ISO4-H5 genome for the use of methylamines. One of the more interesting features of the genomes of members of the Methanomassiliicoccales is the requirement of pyrrolysine for methylamine utilisation (Borrel *et al.* 2014). The ISO4-H5 methylamine:corrinoid methyltransferase genes *mtmB1*, *mtmB2*, *mtbB* and *mttB* all require amber codon read-through to be fully transcribed and incorporate pyrrolysine. The methanol corrinoid protein gene *mtaC2* also requires amber codon read-through to function.

ISO4-H5 encodes four methylcobalamin:CoM methyltransferase *mtaA/mtmA* genes which allow for the use of either methanol and methylamines as substrates for methanogenesis. Experimental evidence is required to assess the substrate specificity, such as conducted in *M. acetivorans* (Bose *et al.* 2006; Bose *et al.* 2008). The methylthiol:CoM methyltransferase, MtsA, required for the utilisation of methylthiol as a substrate, has been identified as a bifunctional enzyme capable of transferring the methyl group from substrates onto a coenzyme corrinoid protein, MtsB, as well as transferring the methyl group onto CoM (Tallant *et al.* 2001). While methyltransferase systems for methanol and methylamines have three components, the methylthiol methyltransferase only requires the MtsA itself and the corresponding corrinoid protein MtsB (Tallant *et al.* 2001). The methylthiol utilisation genes *mtsA* and *mtsB* in ISO4-H5 are homologous to enzymes in *Methanosarcina barkeri* (Tallant and Krzycki 1997), and one of the substrates of the ISO4-H5 MtsA, methyl-3-methylthiopropionate, has been experimentally verified in Chapter 5.

ISO4-H5 is predicted to encode an alcohol dehydrogenase (AR505\_0483), and ethanol is known to be a methanogenic substrate for methanogens such as *Methanofollis ethanolicus* (Imachi *et al.* 2009). Ethanol has also been observed to enhance methanogenesis in *Mbb. ruminantium* M1<sup>T</sup>, presumably by the oxidation of ethanol by its NADP-dependent alcohol dehydrogenases and capture of the reducing potential as F<sub>420</sub>H<sub>2</sub> by the action of their NADPH-dependent F<sub>420</sub> dehydrogenase (Leahy *et al.* 2010). This metabolism is predicted to spare some of the H<sub>2</sub> or formate normally used to produce F<sub>420</sub>H<sub>2</sub> and thus stimulate growth (Leahy *et al.* 2010). Since neither cofactor F<sub>420</sub> biosynthesis nor NADPH-dependent F<sub>420</sub> dehydrogenase genes were predicted within the ISO4-H5 genome, ISO4-H5 is unlikely to utilise ethanol as a supply of reducing potential for methanogenesis. The role of the alcohol dehydrogenase in ISO4-H5 ethanol metabolism requires additional experimental investigation.

For all methanogenic substrates, the methyl group is attached to CoM and reduced by methyl-CoM reductase to produce methane (Thauer 1998). Two different isozymes of the methyl-CoM reductase have been observed in nature; MCR I, abbreviated as Mcr, and MCR II, abbreviated as Mrt. The thermophilic methanogen, *Methanobacterium thermoautotrophicum* harbours both isozymes and appear to be differentially regulated at different stages of growth (Rospert *et al.* 1990; Bonacker *et al.* 1993; Pihl *et al.* 1994; Reeve *et al.* 1997; Friedrich 2005). The two isozymes display different operon structures within a genome, *mcrBDCGA* and *mrtBDGA*. The gene organization of the methyl-CoM reductase in ISO4-H5 suggests it belongs to the MCRII/Mrt isozyme, similar to what was found in other strains of Methanomassiliicoccales

(Lang *et al.* 2015). The metagenome and metatranscriptome has been characterised from rumen contents of high and low methane emitting sheep, it was found the increased methane emission was primarily associated with increases in the expression of *mcr/mrt* operon (Shi *et al.* 2014). As mentioned above, some methanogens harbour both the Mcr and Mrt, such as the *Methanobrevibacter* sp. D5 described in Chapter 6, while some *Methanobrevibacter*s spp. only possess Mcr, as described in *Mbb. ruminantium* M1<sup>T</sup> (Leahy *et al.* 2010). The *mcr/mrt* operon were classified into three groups, and high expression of two groups were associated with increased methane emission, the group extended from the *mcr* operon of *Mbb.* spp, and the group that extended from the *mrt* operon os ISO4-H5 (Shi *et al.* 2014). This suggests ISO4-H5 and similar methanogens play an important role in conferring methane emission of the rumen (Shi *et al.* 2014), which must be taken into consideration when formulating strategies to lower methane emission.

The catalytic action of Mrt depends on the corphin, cofactor F<sub>430</sub>, a the nickel-containing prosthetic group in the active site of methyl-coenzyme M reductase (Scheller *et al.* 2010). The biosynthetic pathway for cofactor F<sub>430</sub> is proposed to involve seven steps, but only the enzyme involved in the first step, uroporphyrin-III C-methyltransferase, CorA, has been experimentally proven (Pfaltz *et al.* 1987; Thauer and Bonacker 1994). The enzymes involved in the remaining steps of the pathway have yet to be identified (Pfaltz *et al.* 1987). ISO4-H5 is predicted to encode a *corA* gene (AR505\_1041), which is homologous to the *corA* gene of *Methanobacterium ivanovii* sharing 43.3% aa identity (Blanche *et al.* 1991). To date, the cofactor F<sub>430</sub> has been uniquely conserved across methanogens (Diekert *et al.* 1981), therefore it is likely ISO4-H5 also possess cofactor F<sub>430</sub>.

CoM is required for the methanogenesis pathway as all methyl groups are transferred to CoM before being reduced to methane (Thauer *et al.* 2008). The ISO4-H5 genome does not contain the *comABCDE* genes required for CoM biosynthesis (Graupner *et al.* 2000). An alternative method of CoM biosynthesis utilising sulfopyruvate biosynthesis has been proposed, involving the genes, threonine synthase-like L-cysteate synthase and an aspartate aminotransferase (Graham *et al.* 2009). Unlike *Methanosarcina acetivorans* whose cysteate synthase was discovered originally as threonine synthase (Graham *et al.* 2009), only one threonine synthase is present within the ISO4-H5 genome, and as such, is unlikely to be a L-cysteate synthase. This means both known pathways of CoM biosynthesis are predicted to be absent in the ISO4-H5 genome, suggesting ISO4-H5 likely requires an external supply of CoM to survive. This is not uncommon within the rumen environment, *Mbb. ruminantium* M1<sup>T</sup> also requires external

CoM to survive (Leahy *et al.* 2010). The lack of CoM biosynthesis genes is consistent with the original enrichment condition of ISO4-H5, which included 10 mM CoM in the medium (Jeyanathan, 2010), and CoM has been used routinely for in vitro cultivation of this methanogen.

The heterodisulfide reductase Hdr is essential for cofactor regeneration during methanogenesis. Hdr has been observed in two classes; a cytoplasmic HdrABC complex that is found across many methanogens, and a transmembrane HdrED found in *Methanosarcina* and *Archaeoglobus* (Mander *et al.* 2004; Buan and Metcalf 2010). HdrE is a membrane-anchored cytochrome, and HdrD contains motifs for iron sulphur cluster binding and two conserved cysteine motifs important in reoxidation of CoM-S-S-CoB (Kaster *et al.* 2011). ISO4-H5 encodes the cytoplasmic HdrABC genes, as well as two HdrD genes without a HdrE gene counterpart. In *Methanosarcina*, HdrDE is coupled to a  $F_{420}H_2$  dehydrogenase (Fpo) complex during methylotrophic methanogenesis (Baumer *et al.* 2000). In order for Hdr to function and recycle the CoB and CoM cofactors, it requires electron donors. In other methanogens the electron donor is reduced cofactor  $F_{420}$ , which is reduced by the  $F_{420}$  reducing hydrogenase (Frh), with hydrogen being the electron donor (Kulkarni *et al.* 2009; Thauer *et al.* 2010). However, ISO4-H5 lacks the genes encoding Frh. Morphological evidence suggests the ISO4-H5 cell lacks cofactor  $F_{420}$ , as it does not possess the typical fluorescence emitted by cofactor  $F_{420}$  under 420 nm (Jeyanathan 2010). *Methanomassiliicoccus luminyensis* was initially reported to be autofluorescent (Dridi *et al.* 2012) but later reported as non-fluorescent (Lang *et al.* 2015), while both strains of Methanomassiliicoccales isolated from the termite and millipede gut have been reported to lack autofluorescence (Paul *et al.* 2012). The methyl viologen hydrogenase, Mvh, has been known to utilise  $H_2$  and supply electrons to the Hdr complex using ferredoxin (Thauer *et al.* 2010). Therefore, it is predicted that ISO4-H5 acquires electrons via Mvh.

The Fpo enzyme is a large transmembrane complex that can oxidize  $F_{420}H_2$  and pump protons out of the cell (Deppenmeier 2004). It is highly similar to the prokaryotic NADH:quinone oxidoreductase and the subunits have been named accordingly. It is comprised of 13 subunits, FpoA,B,C,D,F,H,I,J,K,L,M,N,O, and all the genes encoding these subunits, apart from *fpoF*, are within an operon (Deppenmeier 2004). The *fpoF* and *fpoO* genes are absent in the ISO4-H5 genome. FpoF in other methanogens contains a FAD binding site and an iron-sulfur cluster binding motif, and is thought to act as a receptor of  $F_{420}H_2$  (Welte and Deppenmeier 2011). FpoO also contains an iron-sulfur cluster binding motif, and is likely responsible for reduction

of methanophenazine. ISO4-H5 is predicted to lack both methanophenazine and cofactor F<sub>420</sub>, which would make *fpoF* and *fpoO* genes unnecessary, and this may explain the absence of these genes from the ISO4-H5 genome. The absence of cofactor F<sub>420</sub> from ISO4-H5 means another electron carrier must be used. *Methanosaeta* spp. also lack the *fpoF* gene, and it has been proposed that a ferredoxin can act as an alternative electron carrier (Welte and Deppenmeier 2011). Ferredoxin has been observed to couple Hdr and Mvh electron bifurcation between methanogenesis and energy generation (Kaster *et al.* 2011). ISO4-H5 is predicted to encode three ferredoxin genes. In another member of Methanomassiliicoccales, *Candidatus Methanoplasma termitum*, the HdrD that contains an iron-sulfur binding cluster has been proposed to physically interact with the Fpo-like complex forming an energy-converting ferredoxin:heterodisulfide oxidoreductase (Lang *et al.* 2015). The electron bifurcation proposed to occur at the “*Ca. Methanoplasma termitum*” Mvh/HdrABC complex means that only the electrons of every second hydrogen oxidized by the Mvh/HdrABC complex will go to the Fpo-like/HdrD complex. This will reduce the ATP production and growth yield of this organism, but will mean its threshold for hydrogen will be lower (Lang *et al.* 2015). The energy conservation system in ISO-H5 is predicted to be identical to “*Candidatus Methanoplasma termitum*” and this has several consequences for its existence in the rumen. ISO4-H5 is likely to have a low ATP yield and its growth will therefore be slower. However, ISO4-H5 is also predicted to have a lower hydrogen threshold which may offer it a competitive advantage against other rumen methylotrophic methanogens. *Methanosphaera stadtmanae* for example, has the same general metabolic stoichiometry for methane formation from methanol, but pumps two ions per methane formed (Thauer *et al.* 2008) and is likely to grow faster, but with a higher hydrogen threshold. While the hydrogenotrophic and methylotrophic methanogens may undergo cycles of growth and decline as ruminal hydrogen levels rise and fall, ISO4-H5 and other rumen Methanomassiliicoccales are likely to grow more consistently, albeit at a slower rate.

A number of protein-coding genes have been found as genetic markers for methanogenic archaea. The function of these proteins remain unknown but they are hypothesised to be methanogenesis related, and as such have been termed methanogenesis marker proteins (MMP) (Gao and Gupta 2007). There are 17 MMPs currently described, all of which are conserved among methanogens (Gao and Gupta 2007). ISO4-H5 lacks MMPs 9, 10, 12, 14. It is possible these four MMPs are associated with parts of the methanogenesis pathways that are absent in ISO4-H5.

The central carbon metabolism pathway of ISO4-H5 is depicted in detail in Figure 3.7. The genes present suggest acetate is the primary carbon source of ISO4-H5, and this is consistent with the requirement of acetate for growth (Jeyanathan, 2010). In methanogens, one carbon metabolism begins with the conversion of NADPH to formate by formate dehydrogenase (Andreesen and Ljungdahl 1974), and the incorporation of formate into tetrahydrofolate via formate-tetrahydrofolate ligase. Downstream reactions lead to the production of methyl-tetrahydrofolate, which then condenses with carbon monoxide to form acetyl-CoA by the carbon monoxide dehydrogenase complex (Stupperich *et al.* 1983). Acetyl-CoA can enter into the citric acid cycle or be converted to pyruvate by pyruvate synthase and enter gluconeogenesis. ISO4-H5 utilises acetate as the main carbon source via acetyl-CoA synthetase. The acetyl-CoA synthetase appears to be a single, fused enzyme instead of two subunits commonly found in other organisms, for example within the archaeal genome of *Pyrococcus furiosus* (Musfeldt *et al.* 1999). The ISO4-H5 genome is predicted to lack a methylenetetrahydrofolate reductase, a methyltetrahydrofolate:corrinoid methyltransferase and a carbon monoxide dehydrogenase complex, which suggests ISO4-H5 is unable to produce acetyl-CoA from formate, unlike B10 and Mx1 (Lang *et al.* 2015). Although, ISO4-H5 can utilise formate to produce 5,10 methylene-tetrahydrofolate via the formate-tetrahydrofolate ligase (*fhs*) and bifunctional NADP-dependent methylene tetrahydrofolate dehydrogenase (*fold*) genes, it is not predicted to incorporate carbon monoxide. The 5,10 methylene-tetrahydrofolate can be used for pyrimidine salvage and biosynthesis via the thymidylate synthase gene. It is interesting to note that in an environment of limited acetate, other Methanomassiliicoccales, and indeed *Methanosphaera* spp., may outcompete ISO4-H5 due to their ability to use formate as a carbon source. However, acetate is usually abundant in the rumen and the acetate requirement of ISO4-H5 may reflect its adaptation to this environment.

During central metabolism acetyl-CoA may proceed into the TCA cycle or gluconeogenesis pathway. Genomes of members of the Methanomassiliicoccales have been reported to possess an incomplete TCA cycle (Lang *et al.* 2015). The incomplete TCA cycle of ISO4-H5 begins with fumarate produced from arginine biosynthesis and ends with 2-oxoglutarate, which feeds into glutamate biosynthesis (Pitson *et al.* 1999). Fumarate is one of the endpoints of the TCA cycle, it is recycled into purine ribonucleotide by PurB (Zhang *et al.* 2008). The gluconeogenesis pathway paves the way to the pentose phosphate pathway. The pentose phosphate pathway connects glyceraldehyde-3-phosphate from gluconeogenesis to 5-phospho-D-ribose 1-diphosphate (PRPP) for purine biosynthesis (Sprenger 1995; Kato *et al.* 2006). The

ISO4-H5 pathway is predicted to rely heavily on the broad substrate specificity of the enzymes transketolase and transaldolase (Cunin *et al.* 1986). The additional D-xylulose-5-phosphates generated by the process can be recycled by Rps, HxlA, HxlB back to D-fructose-6-phosphate (Kato *et al.* 2006). The *hxlA* and *hxlB* genes do not form an operon, unlike in *Bacillus subtilis* (Yasueda *et al.* 1999).

The ISO4-H5 genome is predicted to contain all the genes required to synthesize 19 amino acids, with the exception of glycine and methionine. Therefore ISO4-H5 may require an external supply of glycine, but a transporter for glycine has yet to be identified within the genome of ISO4-H5. The methionine biosynthesis pathway is depicted in Figure 3.9. Several genes required for the synthesis of methionine are not found in the genome of ISO4-H5 and whether ISO4-H5 is capable of methionine biosynthesis remains unclear. There are two potential routes of methionine biosynthesis; a three step route that uses *O*-acetyl-L-homoserine as intermediate lacks the *met2* gene in second step, and a five step route that uses *O*-succinyl-L-homoserine as an intermediate lacks all the enzymes from the third step onwards (Figure 3.09). The first step is universally carried out by MetL (Belfaiza *et al.* 1984) and ISO4-H5 encodes a *metL* gene. This gene is predicted to encode a homoserine dehydrogenase domain at its C-terminal end (Belfaiza *et al.* 1984). MetL from *E. coli* is a bifunctional enzyme with the N-terminal end acting as an aspartate kinase (Belfaiza *et al.* 1984). Without the N-terminal aspartate kinase domain in the the ISO4-H5 *metL* gene, it is predicted to carry out only one function, that being homoserine biosynthesis from aspartate-semialdehyde. The methionine biosynthesis pathway branches at L-homoserine. The Met2, homoserine *O*-acetyltransferase, is involved in producing *O*-acetyl-L-homoserine (Yamagata 1987), and ISO4-H5 lacks a *met2* gene. The gene *O*-acetyl-L-homoserine sulphydrylase (AR505\_0738) is also known as *O*-acetylhomoserine aminocarboxypropyltransferase, Met17. This gene has been observed to convert the *O*-acetyl-L-homoserine to L-methionine using methanethiol in *Neurospora* (Moore and Thompson 1967). It has been shown to react with other thiols to form homocysteine, an intermediate for methionine biosynthesis in many organisms (Shimizu *et al.* 2001). There are four enzymes capable of converting homocysteine to L-methionine, MetH, MetE, Met6 and Bhmt (Figure 3.09), but none of these genes have homologues within the ISO4-H5 genome. In the other route, ISO4-H5 is predicted to encode MetA that can produce *O*-succinyl-L-homoserine but there are no enzymes encoded in the genome to process this substrate further. The enzymes present in both routes could complement each other to produce methionine if the Met17 gene could utilise *O*-succinyl-L-homoserine as a substrate instead of *O*-acetyl-L-

homoserine. Methionine is also an important part of the *S*-adenosyl-L-methionine (SAM) cycle (Loenen 2006) and SAM is an important cofactor involved in methylation reactions. Homologues of enzymes required for the SAM cycle and SAM dependent DNA methylation are predicted within the ISO4-H5 genome, apart from methionine synthase (Table 3.12). ISO4-H5 is unable to produce spermidine from SAM due to the lack of a spermidine synthase gene. It is possible that ISO4-H5 has adapted a novel mechanism of L-methionine production, as it is unlikely for an organism to carry incomplete versions of both biosynthesis and utilisation of methionine, as well as the SAM cycle, if they are not being utilised in some way by the organism.

An interesting aspect of ISO4-H5 amino acid biosynthesis was that although it is predicted to be able to produce asparagine, the amino source of the enzymatic reaction remains elusive. The asparagine synthase B AsnB converts aspartate to asparagine (Scofield *et al.* 1990). The ISO4-H5 genome is predicted to encode two AsnB genes (predicted molecular weight, 35kD and 28kD), which are smaller than the experimentally verified *E. coli* AsnB (63kD) (Scofield *et al.* 1990). A multiple alignment reveals that the ISO4-H5 *asnB* genes only possess the C-terminal domain that is responsible for binding of ATP and aspartate, but not the N-terminal domain that hydrolyses the glutamine (Larsen *et al.* 1999). This suggests the ISO4-H5 AsnB does not utilise glutamine as the amino source. Further study on the enzyme kinetics of this AsnB would be required to identify the amino source.

As mentioned earlier, the amino acid pyrrolysine is required for expression of enzymes involved in methylamine utilisation. Pyrrolysine was discovered as the 22<sup>nd</sup> amino acid in 2002 in the crystal structure of monomethylamine methyltransferase MtmB from *Methanosarcina barkeri* (Hao *et al.* 2002). Pyrrolysine is encoded by the UAG codon, also known as the amber codon (Hao *et al.* 2002). The pyrrolysine usage extends beyond the order Methanomassiliicoccales and is found amongst other anaerobic archaea as well as bacteria (Herring *et al.* 2007; Prat *et al.* 2012). Methanogenic archaea that utilise pyrrolysine include *Methanosarcina acetivorans* (Gaston *et al.* 2011), *Methanococcoides burtonii* (Gaston *et al.* 2011), *Methanohalobium evestigatum* (Gaston *et al.* 2011), and *Methanohalophilus mahii* (Gaston *et al.* 2011). Four genes are required for the biosynthesis of pyrrolysine from lysine, as well as for the biosynthesis of the pyrrolysine-tRNA (Table 3.12), *pylS*, *pylB*, *pylC*, *pylD* (AR505\_1325 - 1322). The amber suppressor tRNA(CUA) gene, *pylT* is also required. The organisation of the pyrrolysine biosynthesis operon in ISO4-H5 (Figure 3.06) is identical to what has been found in the *Methanosarcina barkeri* genome (Zhang *et al.* 2005) as well as



being highly similar to what has been described for other members of the Methanomassilicoccales order (Srinivasan *et al.* 2002; Borrel *et al.* 2014).

The recoding of the amber codon for pyrrolysine is not the only alternative coding of a stop codon observed in nature. The opal stop codon UGA has been observed to encode selenocysteine in many organisms (Johansson *et al.* 2005), and tryptophan has been observed to be encoded by UGA in *Mycoplasma capricolum* (Yamao *et al.* 1985). Recent advancements in single cell sequencing technology has acquired genome sequences of uncultivated micro-organisms, and has also revealed the recoding of UGA to glycine in *Gracilibacteria* (Rinke *et al.* 2013). In ISO4-H5 the opal codon acts as the major stop codon and accounts for 84.1% of all predicted stop codons, while the ochre stop codon accounts for 13.4% (Table 3.5). While read-through of the amber codon may be essential for the complete translation of functional methylamine methyltransferase(s), it can also produce effectively a ‘nonstop’ mutation to other proteins carrying the in-frame amber codon (Wong and Schwartzbach 2015). The ‘nonstop’ mutation has been studied in humans, where 119 identified ‘nonstop’ mutations in 87 proteins were found to be associated with an inherited disease, as the extended amino acid often caused peptides to fold differently and caused alteration or loss of functionality. The longer the peptide extension the more likely the resulting peptide was found to fold differently and lose the original function (Ameri *et al.* 2007; Hamby *et al.* 2011).

The recoding of stop codons leads to an interesting point on how ISO4-H5 regulates pyrrolysine incorporation. Based on sequence homology, there are 10 genes predicted to utilise the amber codon as a stop codon rather than pyrrolysine incorporation (Table 3.8). This includes nine IS elements and one adenylate kinase gene. The importance of pyrrolysine to some of these genes is unclear, however it is obvious that without pyrrolysine, ISO4-H5 would be unable to utilise methylamines as a source of methyl-compounds for methanogenesis. The abundance of methylamines in the rumen environment (Saleem *et al.* 2013) combined with the low number of genes utilising the amber codon as a stop codon, (including the uncertain genes only accounts for 1.5% of all genes), suggests that the pyrrolysine operon is constitutively expressed and no regulation mechanism is in place within ISO4-H5. Therefore, the additional 10 genes predicted to use the amber codon as a stop codon would encode extended and likely misfolded, non-functional proteins. The loss of function to the nine IS elements means the nine IS elements may extend and potential loss function, however the inverted repeats flanking the IS element would not be influenced by transcription itself, and the inverted repeats may yet potentially influence the expression of surrounding genes. With regard to the adenylate kinase

gene, if the 20 extended amino acids caused the adenylate kinase to misfold, ISO4-H5 does carry a second copy of an adenylate kinase (AR505\_0262) which does not have a pyrrolysine codon. ISO4-H5 may lose some energy through producing a non-functional adenylate kinase, but it would not lose this cellular function completely.

The ISO4-H5 is predicted to possess a near complete pathway of *de novo* nucleotide biosynthesis, as illustrated in Figure 3.12, and the purine biosynthesis pathway is similar to what has been reported for other members of Methanomassiliicoccales (Borrel *et al.* 2014).. The enzyme required for conversion of IMP to XMP remains to be identified. One of the steps, the conversion of amino-(5-phospho-ribosyl) imidazole (AIR) to amino-(5-phospho-ribosyl) imidazole-4-carboxylate (CAIR) is helpful to ISO4-H5 energy conservation. Instead of the ATP-dependent two-step reaction catalysed by carboxyaminoimidazole ribonucleotide synthetase (PurK) and carboxyaminoimidazole ribonucleotide mutase (PurE), AIR is converted to CAIR by phosphoribosylaminoimidazole carboxylase (Ade2) which incorporates CO<sub>2</sub> (Stotz and Linder 1990). This reduces the ATP requirement of purine *de novo* biosynthesis in ISO4-H5 and suggests ISO4-H5 may have a preference to enzymes that do not consume ATP.

The predicted *deoA*, *rbcL* and *e2b2* operon is an interesting feature of the ISO4-H5 genome involved in purine degradation and carbon fixation. DeoA recycles AMP to adenine and ribose-1,5-bisphosphate, which can be isomerized to ribulose-1,5-bisphosphate, which is supplied to RubisCO for CO<sub>2</sub> fixation as well as re-entry into the pentose phosphate pathway via 3-phospho-D-glycerate (Watson *et al.* 1999; Sato *et al.* 2007). ISO4-H5 lacks a detectable phosphoenolpyruvate carboxylase and the hydrogen carbonate is likely not utilised for oxaloacetate production (Smith and Ferry 1999). This allows ISO4-H5 to recycle nucleotides as well as acquire an additional carbon source. However, the carbon fixation via RubisCO may not sustain the cell as the sole carbon source.

The archaeal phospholipids differentiate themselves from phospholipids from other domains of life, by being composed of isoprenoid chains and glycerol-1-phosphate, joined by ether linkages (Nishimura and Eguchi 2006). The predicted ISO4-H5 lipid biosynthesis pathway is illustrated in Figure 3.11, and shows that the biosynthesis of isoprenoids from the mevalonate pathway is similar to what was previously established in *Methanocaldococcus jannaschii* (Kuzuyama 2002; Grochowski *et al.* 2006). The ISO4-H5 genome is predicted to lack phosphomevalonate decarboxylase, MvaD, however it has been proposed that conversion of mevalonate-5-phosphate to isopentenyl phosphate could be carried out by an unidentified ATP-

independent decarboxylase (Vinokur *et al.* 2014). In methanogens such as *Methanothermobacter thermoautotrophicus* and *Mbb. ruminantium* M1<sup>T</sup>, the CDP-archaeol proceeds towards production of archaetidylserine and archaetidylethanolamine (Morii *et al.* 2000; Morii and Koga 2003; Leahy *et al.* 2010), which ISO4-H5 is unable to produce due to lack of an archaetidylserine synthase and decarboxylase. Instead, ISO4-H5 is predicted to produce cardiolipin, a phospholipid identified within the inner mitochondrial membranes as well as in bacterial cell envelopes (Schlame 2008). It has been proposed to function as a proton trap by bicyclic resonance between phosphates and a central hydroxyl group, capable of protein binding with chaperone-like activity and stabilising the curved region of a phospholipid membrane (Haines and Dencher 2002; Schlame and Ren 2009). The properties of cardiolipin may be beneficial to the survival of ISO4-H5 within the rumen, and contribute to its unusual cell morphology. ISO4-H5 is devoid of cell wall protection, therefore the presence of cardiolipin may make the ISO4-H5 cell envelope more flexible, although it remains physically fragile. In exposure to physical stresses, it is predicted that methanogens equipped with a pseudomurein and S-layer-containing cell envelopes would have a better chance of survival than ISO4-H5-like organisms.

In methanogens, the cell envelope often consists of pseudomurein polysaccharides enclosed in a S-layer (Arbing *et al.* 2012). The absence of UDP-*N*-acetylglucosamine diphosphorylase, glucosamine-1-phosphate *N*-acetyltransferase and glucosamine-fructose-6-phosphate aminotransferase suggests the ISO4-H5 genome does not harbour sufficient genes to produce a functional pseudomurein. The S-layer of *Aneurinibacillus thermoaerophilus* DSM 10155 is found to be composed of crosslinking rhamnose and glycerol-manno-heptose (Wugeditsch *et al.* 1999). ISO4-H5 is predicted to encode a dTDP-L-rhamnose biosynthesis (AR505\_0552 – 0554) gene cluster and glycerol-manno-heptose biosynthesis genes (AR505\_1754, AR505\_1770 – 1772), which suggests ISO4-H5 may be capable of producing an S-layer-like extracellular structure.

The predicted extracellular proteins of ISO4-H5 were analysed to gain insight into the interaction of ISO4-H5 with the rumen environment. Proteins require a signal sequence in order for a cell to transport the protein to the cell envelope or to export the protein outside the cell (Eichler and Moll 2001). ISO4-H5 is predicted to use the SRP and Sec61 translocation complex for translocation of gene products (Zwieb and Bhuiyan 2010). There are 101 ISO4-H5 genes predicted to encode an N-terminal signal peptide (Figure 3.04), 35 proteins are predicted with two or more transmembrane domain, and an additional 12 proteins are predicted to be anchored

through a C-terminal transmembrane domain. These genes include 10 ABC transporter family proteins, four transporters involved in drug resistance, a mechanosensitive ion channel (AR505\_0163) and a Na<sup>+</sup>/H<sup>+</sup> antiporter (AR505\_0385). There are 39 proteins predicted to be anchored through an N-terminal transmembrane domain. This includes four ABC transporter family proteins, two of which are classified as regulatory genes; a carbon starvation protein, CstA (PF02554) (Schultz and Matin 1991), anchored with 12 transmembrane domains and a secreted sugar fermentation stimulation protein, SfsA (PF03749) (Kawamukai *et al.* 1991). The CstA may sense the environment and regulate ISO4-H5 carbon metabolism. It is unclear how a DNA-binding SfsA would function extracellularly. There are 45 adhesin-like proteins predicted to be surface-located proteins. In the *Methanobrevibacter smithii* genome, adhesin-like proteins were associated with host mucosal glycosaminoglycans (Samuel *et al.* 2007). Also, physical co-localization of *Methanobrevibacter ruminantium* M1<sup>T</sup> with *Butyrivibrio proteoclasticus* has been previously correlated with up-regulation of adhesin-like proteins (Leahy *et al.* 2010; Ng *et al.* 2016). Domain repeats within adhesins are often involved in binding (Cabanes *et al.* 2002). ISO4-H5 encodes 19 adhesin-like proteins predicted to harbour domain repeats (Table 3.10). One adhesin-like protein is predicted to be involved in binding to a dockerin of a cellulosome (PF00963) (Shimon *et al.* 1997). Cellulosomes are cell surface protrusions produced in some cellulose hydrolysing microbes (Lamed *et al.* 1987), which is also present in the rumen (Oyeleke and Okusanmi 2008), including the hydrogen producing *Ruminococcus flavefaciens* (Kirby *et al.* 1997; Ding *et al.* 2001; Jindou *et al.* 2008). The binding of ISO4-H5 to a cellulosome could assist in acquiring hydrogen required for methanogenesis. A particular type of high molecular weight adhesins have been identified in *Haemophilus influenza* with a role in the attachment to human epithelial cells (Barenkamp 1996; Barenkamp and St Geme 1996). ISO4-H5 contains seven adhesin-like proteins larger than 200kDa and it is possible these genes could facilitate interaction of ISO4-H5 with its ovine host or with other rumen microorganisms. ISO4-H5 was originally enriched from the ovine rumen with the bacterium *Succinivibrio dextrinosolvens*, and it has been demonstrated recently that there is a strong association between the order Methanomassiliicoccales and *Succinivibrio* spp. (Henderson *et al.* 2015). Whether or not ISO4-H5 uses its adhesins to associate with particular rumen surfaces or microbes remains unclear.

ISO4-H5 is predicted to utilise ammonium from the environment as a nitrogen source via either GdhA or GlnA and an ammonia transporter. Glutamate and glutamine are central to nitrogen metabolism as well as representing the protein building blocks for the biosynthesis of other

amino acids. ISO4-H5 also utilises GdhA for interconversion between 2-oxoglutarate and glutamate with ammonia and  $\text{NAD(P)}^+$ , as well as type-III GlnA for ATP driven interconversion of glutamate to glutamine (Helling 1998) (Table 3.11). The predicted nitrogenase and nitroreductase genes along with molybdenum cofactor biosynthesis genes suggests ISO4-H5 may be capable of nitrogen fixation (Leigh 2000). However nitrogen fixation ability cannot be concluded based on the genes alone, as it has been shown in *Methanocaldococcus jannaschii* that these genes does not necessary confer nitrogen fixation activity (Staples *et al.* 2007).

Nicotinamide adenine dinucleotide (NAD) is an electron carrier involved in redox reactions (Pollak *et al.* 2007) and is an important coenzyme in methanogens, as demonstrated by *Mbb. ruminantium* M1, which utilises NADP to relay reducing potential from ethanol to  $\text{F}_{420}$  (Leahy *et al.* 2010). ISO4-H5 does not possess the full pathway for *de novo* NAD biosynthesis, as the *nadB* and *nadD* genes are absent. The L-aspartate dehydrogenase AspDH has been reported to be able to replace NadB (Yang *et al.* 2003), the L-aspartate dehydrogenase gene is predicted in the ISO4-H5, which may function similary in ISO4-H5. ISO4-H5 is predicted to carry full set of genes required to salvage nicotinamide and nicotinic acid for NAD biosynthesis, so it seems likely that NAD can be synthesised in ISO-H5.

### 3.4 Conclusions

ISO4-H5 has a genome similar to the those of other members of the order Methanomassiliicoccales, Mx1201, B10 and BRNA1. ISO4-H5 encodes the key genes and pathways required for hydrogen-dependent methylotrophic methanogenesis by reduction of methyl substrates, without the ability to oxidize methyl substrates to carbon dioxide. The wide range of methyl substrates predicted to be used by ISO4-H5 suggests it is more metabolically versatile than other methylotrophic methanogens within the rumen. Members of Methanomassiliicoccales co-exist in the rumen with *Methanosphaera* spp. (Janssen and Kirs 2008; Seedorf *et al.* 2015) and share similar substrate requirements. Methanomassiliicoccales are probably able to outcompete *Methanosphaera* in the rumen at low hydrogen concentrations, due to the lower thresholds conferred by the low ATP gain. However, they are probably disadvantaged when substrate concentrations are high and the low ATP yield limits their ability to proliferate quickly. The variability of fermentation rates in the rumen associated with periods of feeding or fasting is therefore expected to give both groups of methylotrophic methanogens opportunities to grow.

ISO4-H5 appears to be reliant on the Hdr, Mvh and Fpo-like complexes for electron bifurcation, membrane potential generation and energy conservation, as has been described in other members of Methanomassiliicoccales. However, ISO4-H5 is incapable of producing CoM, which suggests that ISO4-H5 has adapted to the rumen environment, where CoM produced by other methanogens would be able to supplement ISO4-H5. ISO4-H5 also lacks the genes encoding cofactor F<sub>420</sub> synthesis, rendering it non-fluorescent under illumination at 420 nm. This trait has also been reported amongst other members of Methanomassiliicoccales, and is likely one of the key characteristics of this particular order of methanogens. However, a culture of B10 has been reported to fluoresce (Arbing *et al.* 2012; Dridi *et al.* 2012; Borrel *et al.* 2013) and this may be consistent with B10 belonging to the deepest branching group within Methanomassiliicoccales (Seedorf *et al.* 2014).

The use of pyrrolysine in proteins carrying out various cellular functions suggests it is important for ISO4-H5. While pyrrolysine is important in methylamine utilisation by all members of Methanomassiliicoccales sequenced thus far, pyrrolysine also appears to play a role in methanol use by ISO4-H5, as the methanol:methyltransferase corrinoid protein, MtaC1, is predicted to contain a pyrrolysine in its full length protein. The use of pyrrolysine and the possession of a Fpo-like complex by ISO4-H5 adds further weight to the hypothesis that the order Methanomassiliicoccales is evolutionary closer to the order Methanosarcinales, supporting findings from a previous phylogenetic study (Janssen and Kirs 2008). By analyzing the genome of ISO4-H5, our knowledge of the order Methanomassiliicoccales has been expanded, and together with the genomes of other members of the Methanomassiliicoccales, will be an important resource for the development of methane abatement technologies in ruminants.

## Chapter 4

### Genomic insight into members of the Methanomassiliicoccales isolated from the rumen as revealed by comparative genome analysis

#### 4.1 Introduction

Members of the order Methanomassiliicoccales are methylotrophic methanogens first detected in the rumen (Yanagita *et al.* 2000) and subsequently in a variety of other anaerobic environments, including landfill (Huang *et al.* 2002), marine hydrothermal vents (Takai and Horikoshi 1999), lake sediments (Nusslein *et al.* 2001), rice paddies (Conrad *et al.* 2006), humus-feeding larva (Egert *et al.* 2003), termite and millipede hindguts (Shinzato *et al.* 1999; Paul *et al.* 2012), the reticulo-rumen compartment of ruminant animals (Yanagita *et al.* 2000; Tajima *et al.* 2001), and in the human gut and subgingival plaque (Dridi *et al.* 2012; Horz *et al.* 2012). Genomes of three members of Methanomassiliicoccales isolated from human faeces have been sequenced, including “*Candidatus* Methanomassiliicoccus intestinalis Mx1-Issoire”, “*Candidatus* Methanomethylophilus alvus Mx1201” and *Methanomassiliicoccus luminyensis* B10 (Borrel *et al.* 2012; Dridi *et al.* 2012; Borrel *et al.* 2013). In addition, the genome sequence of “*Candidatus* Methanoplasma termitum MpT1” enriched from the termite gut has been reported (Lang *et al.* 2015). Currently, four genome sequences are publically available for bovine rumen isolates, they are Thermoplasmatales archaeon BRNA1 (NCBI Reference Sequence NC\_020892.1), “*Candidatus* Methylophilus” sp. 1R26 (Noel *et al.* 2016), Methanomassiliicoccales archaeon RumEn M1 and RumEn M2 (Sollinger *et al.* 2016). Comparative genome analysis of these reported genome sequences has highlighted the lack of genes encoding the hydrogenotrophic methanogenesis pathway, but confirmed the presence of genes required for methylotrophic methanogenesis via direct reduction of methyl-substrates (Borrel *et al.* 2014). A recent comparative study which included MpT1 from the termite gut has confirmed the utilisation of pyrrolysine and the lack of an rRNA operon structure, and has expanded our understanding of energy metabolism by predicting that the Fpo-complex is responsible for membrane potential generation in the order Methanomassiliicoccales (Lang *et al.* 2015). Currently, the Methanomassiliicoccales can be considered as two clades, with one that harbors more gastrointestinal (GIT) isolates and one that harbors more environmental isolates, but neither clades are exclusive to GIT or environmental isolates (Sollinger *et al.* 2016).

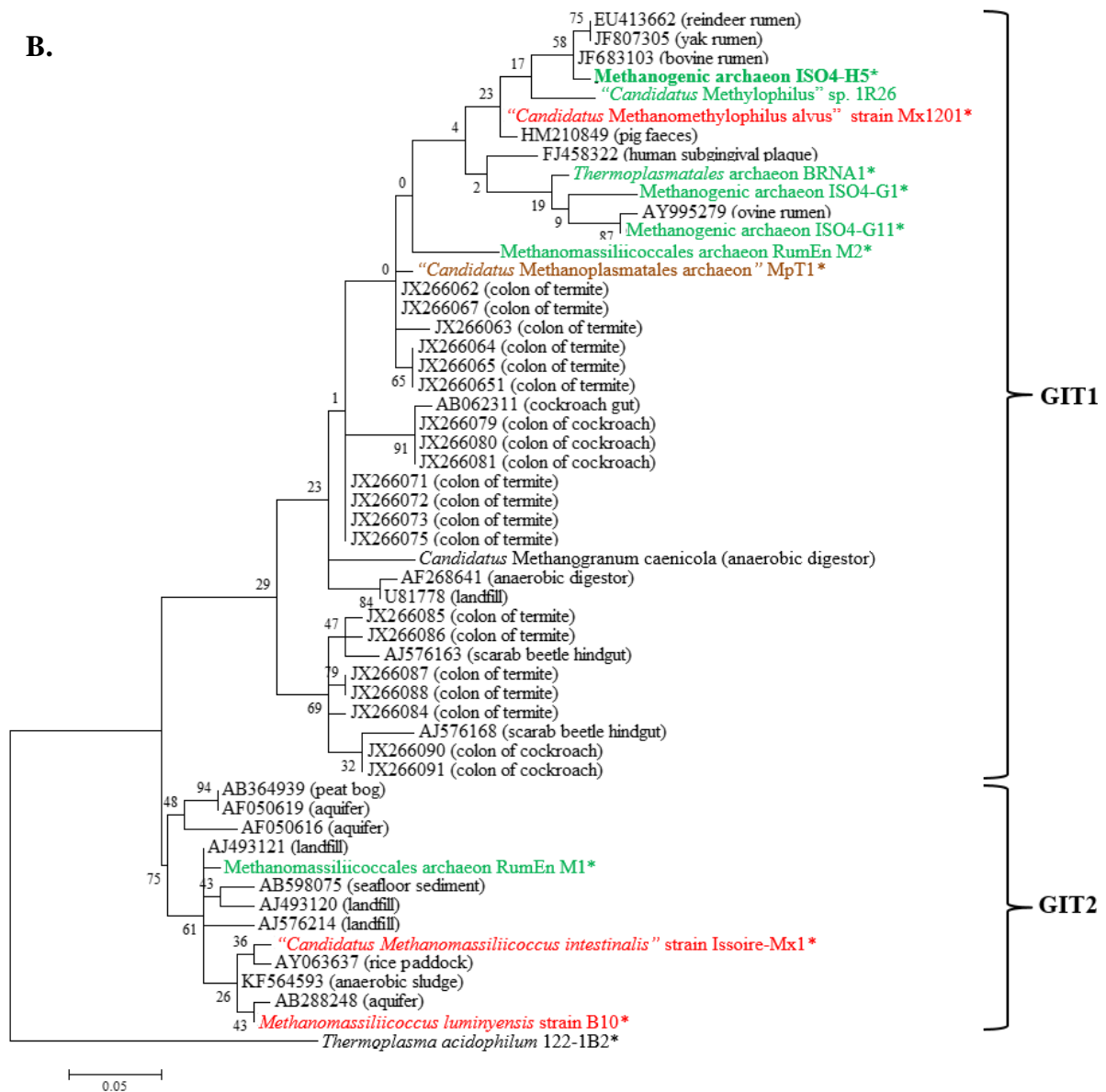
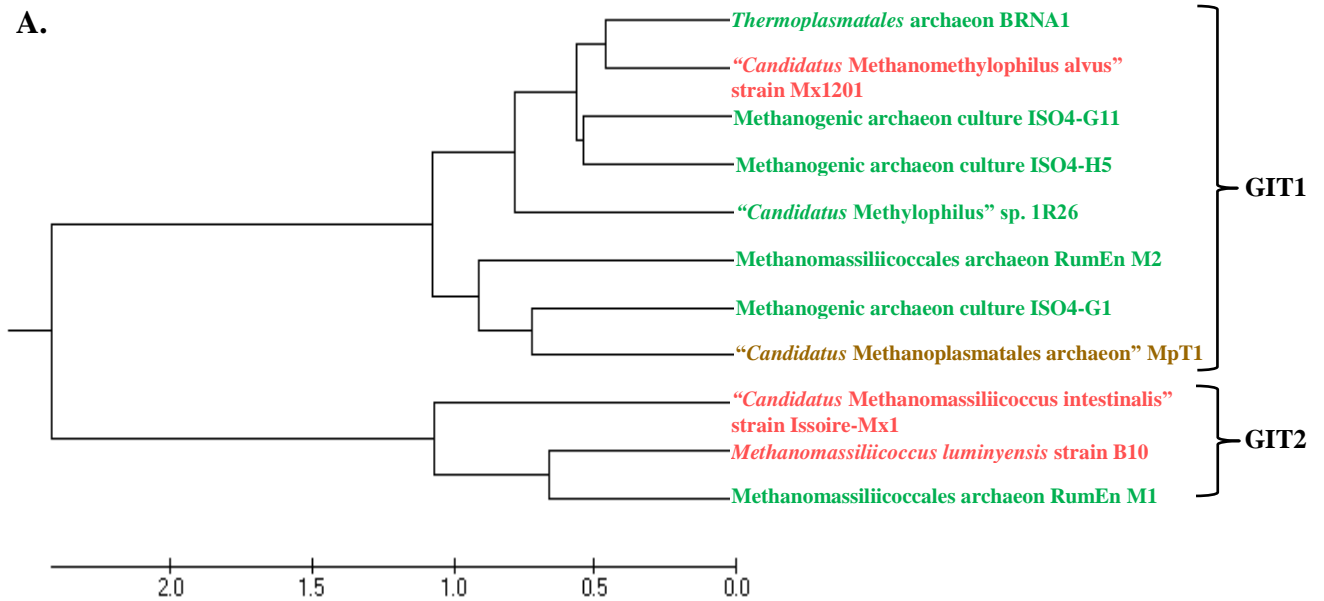
The methanogenic archaeon ISO4-H5, isolated from the ovine rumen, has been sequenced as part of this thesis (see Chapter 3). Furthermore, two strains of the Methanomassiliicoccales order, ISO4-G1 and ISO4-G11, also isolated from the ovine rumen (Jeyanathan 2010) have been sequenced as part of the NZAGRC and PGGRC programmes of research and made available for comparative analysis in this thesis study (Kelly *et al.* 2016). Combined with the genome sequence of Thermoplasmatiales archaeon BRNA1, Methanomassiliicoccales archaeon RumEn M1, RumEn M2 and “*Candidatus Methylophilus*” sp. 1R26 (Noel *et al.* 2016; Sollinger *et al.* 2016), this allows for a detailed investigation into the genomes of members of the rumen Methanomassiliicoccales. In this chapter, I have compared the genome sequences of the rumen Methanomassiliicoccales, with those from other environments. I have focused my analysis on gene conservation to gain insight into the metabolism and physiology of these organisms and have identified specific sets of genes which are important for adaptation to the rumen environment.

## **4.2 Results**

### **4.2.1 Phylogenetic relationship of members of the Methanomassiliicoccales isolated from gut environments**

A FGD tree (Figure 4.1A) was created to assess the similarity between organisms based on the full complement of open reading frames within each genome. Based on this analysis, ISO4-H5 was most closely related to Mx1201, then ISO4-G11 and BRNA1. ISO4-H5 is more distantly related to ISO4-G1 and MpT1 strains. Mx1 and B10 fell within a completely different branch, similar to the 16S rRNA gene based tree (Figure 4.1B). The environmental isolates and gastrointestinal isolates tend to cluster on different branches of the 16S rRNA gene-based tree.





**Figure 4.1 Phylogenetic relationship of members of the order Methanomassiliicoccales. A.** Functional Genome Distribution tree of 11 members of the Methanomassiliicoccales with complete or draft genome sequences. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the tree. **B.** 16S rRNA gene-based phylogenetic tree showing the relationships of methanogenic archaeon ISO4-H5 (shown in bold print) relative to other type and non-type strains within the order Methanomassiliicoccales. \* indicates strain with available genome sequences. The phylogeny was inferred from 16S rRNA gene nucleotide sequences (1,474 bp internal region) aligned using the Maximum Likelihood method based on the Kimura 2-parameter model. Bar: 0.05 substitutions per nucleotide position. The GenBank/nucleotide accession numbers of environmental sequences are displayed with the source habitat given in brackets. GIT1: Gastrointestinal cluster 1, GIT2: Gastrointestinal cluster 2. Strains whose genomes have been sequenced are marked with asterisk. Strains in red text are enriched or isolated from human faeces, strains in green text are enriched or isolated from rumen, and strain in brown text are enriched from termite gut. The initial tree for the heuristic search was obtained by applying the Neighbour-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. All positions containing gaps or missing data were eliminated, giving a total of 455 positions in the final dataset. The 16S rRNA gene sequence from *Thermoplasma acidophilum* 122-1B2 was used as an outgroup.

#### 4.2.2 Genome characteristics

The general genome characteristics of members of the Methanomassiliicoccales are listed in Table 4.1. All of the genomes are predicted to be single, circular chromosomes, with no evidence to support the presence of plasmids.

*Ribosomal, non-coding and transfer RNAs.* ISO4-H5 contains two sets of rRNA genes with an additional 5S rRNA gene, as described in Chapter 3 (Table 3.3). Eight of the Methanomassiliicoccales genomes, including ISO4-G1 and ISO4-G11 have a single set of rRNA genes with an additional 5S rRNA gene, 1R26 has two additional 5S rRNA genes, each of the genes are scattered throughout the genome. MpT1 does not have an additional 5S rRNA gene (Table 4.1). The Methanomassiliicoccales genomes analysed are predicted to carry seven ncRNAs (Table 4.2). Two ncRNAs are conserved: an archaeal RNase P and an archaeal signal recognition particle (SRP) (Table 4.2). Both RNase P and SRP are ribozymes comprised of ncRNA and protein subunits. Two protein subunits of SRP: SRP54 and SRP19 are conserved (Table A.4.17). The RNase P protein components are conserved across eight genomes with protein components 1 and 4, whereas 1R26, Mx1201 and B10 genomes are predicted to only have either protein component 1 or 4. Group II introns have been predicted in the Mx1201, ISO4-H5 and ISO4-G11 genomes. The genomes of ISO4-H5 (AR505\_1776) and ISO4-G11 (ISO4G11\_0442) encode a group II intron maturase. Two predicted group II introns in Mx1201 are located adjacent to an ATP-dependent RNA helicase (MMALV\_11790), the other overlaps with a retron-type RNA-directed DNA polymerase (MMALV\_06720), which is possibly encoded by the group II intron. The second predicted group II intron in the ISO4-H5 genome overlaps with a hypothetical protein (AR505\_0861) without a corresponding maturase. There is only one small nucleolar RNA (snoRNA) predicted in the ISO4-G1 genome, sR41.

**Table 4.1 General genome properties of sequenced members of Methanomassiliicoccales**

	Methanogenic archaeon ISO4-H5	Methanogenic archaeon ISO4-G1	Methanogenic archaeon ISO4-G11	<i>Thermoplasmatales</i> archaeon BRNA1	Methanomassiliicocca les archaeon RumEn M1	Methanomassiliicocca les archaeon RumEn M2	<i>“Candidatus</i> <i>Methylophilus”</i> sp. 1R26
Source	Ovine rumen	Ovine rumen	Ovine rumen	Bovine rumen	Bovine rumen	Bovine rumen	Bovine rumen
Culture status	Isolated	Enriched	Enriched	Enriched	Enriched	Enriched	Enriched
Project status	Complete	Complete	Draft (52 contigs)	Complete	Draft (182 contigs)	Draft (18 contigs)	Draft (50 contigs)
Genome size (bp)	1,937,882	1,593,504	1,901,999*	1,461,105	2,121,026	1,280,797	1,723,106
G+C content (%)	54.0	55.5	56.9	58.3	62.1	54.6	60.4
Number of ORFs	1826	1527	1758	1577	1875	1175	1,465
Coding area (%)	90	92	88	91	68	73.3	65.9
rRNA (5S, 16S, 23S)	3, 2, 2	2,1,1	2,1,1	2, 1, 1	2, 1, 1	2, 1, 1	3, 1, 1
tRNAs (with introns)	47 (3)	47 (3)	47(3)*	47 (3)	46(8)	45(9)	49(3)
Non-coding RNA	8	3	5*	4	3*	2*	2*
Insertion sequences	41	7	29*	7	Nd	Nd	Nd
Prophage	0	0	1	0	0	0	0
CRISPR regions (spacers)	1 (32)	0	0	1 (6)	1 (6)	2 (3, 5)	1 (13)
Adhesin-like proteins	44	18	26*	17			
NCBI accession	CP014214	CP013703	N/A	CP002916.1	LJKK000000000	LJKL000000000	LOPS000000000
Publication	Li <i>et al.</i> , 2016	Kelly <i>et al.</i> , 2016			Sollinger <i>et al.</i> , 2016	Sollinger <i>et al.</i> , 2016	Noel <i>et al.</i> , 2016
	<i>“Candidatus</i> <i>Methanomassiliicoccus</i> <i>intestinalis</i> Mx1- Issoire”	<i>“Candidatus</i> <i>Methanomethylophilu</i> <i>s alvus</i> Mx1201”	<i>Methanomassiliicoccus</i> <i>s luminyensis</i> B10	<i>“Candidatus</i> <i>Methanoplasma</i> <i>termitum</i> MpT1”			
Source	Human faeces	Human faeces	Human faeces	Termite gut			
Culture status	Enriched	Enriched	Isolated	Enriched			
Project status	Complete	Complete	Draft (26 contigs)	Complete			
Genome size (bp)	1,931,651	1,666,795	2,620,233*	1,488,669			
G+C content (%)	41.3	55.6	60.5	49.2			
Number of ORFs	1876	1651	2669	1,415			
Coding area (%)	88	89	86	91			
rRNA (5S, 16S, 23S)	2, 1, 1	2, 1, 1	2, 1, 1	1,1,1			
tRNAs (with introns)	46 (7)	47 (4)	47 (8)*	46 (4)			
Non-coding RNA	3	7	2*	3			
Insertion sequences	16	19	Nd	19			
Prophage	0	0	0	0			
CRISPR regions (spacers)	1 (110)	2 (47, 12)	1 (113)	1 (53)			
Adhesin-like proteins	Nd	Nd	Nd	Nd			
NCBI accession	CP005934.1	CP004049.1	CAJE01000001 - 26	CP010070.1			
Publication	Borrel <i>et al.</i> , 2013	Borrel <i>et al.</i> , 2012	Gorlas <i>et al.</i> , 2012	Lang <i>et al.</i> , 2015			

\*among currently available draft sequences

**Table 4.2 Non-coding RNAs in Methanomassiliicoccales genomes**

	ISO4-H5	ISO4-G1	ISO4-G11	BRNA1	RumEn M1	RumEn M2	1R26	Mx1201	Mx1	B10	MpT1
Archaeal RNaseP (protein components)	+ (1, 4)	+ (1, 4)	+ (1, 4)	+ (1, 4)	+ (1, 4)	+ (1, 4)	+ (1)	+ (4)	+ (1, 4)	+ (4)	+ (1, 4)
Archaeal Signal recognition particle	+	+	+	+	+	+	+	+	+	+	+
Small nucleolar RNA sR41	-	+	-	-	-	-	-	-	-	-	-
Group II catalytic intron	+	-	+	-	-	-	-	+	-	-	-
Group II catalytic intron D1-D4-1	-	-	-	-	-	-	-	+	-	-	-
Group II catalytic intron D1-D4-3	+	-	-	-	-	-	-	+	-	-	-
Group II catalytic intron D1-D4-7	+	-	+	-	-	-	-	-	-	-	-
TPP riboswitch	-	-	-	-	+	-	-	-	-	-	-

+ Indicate the presence of the ncRNA, -indicates absence of the ncRNA. (Number) indicates the protein subunits predicted to be associated with the ncRNA.

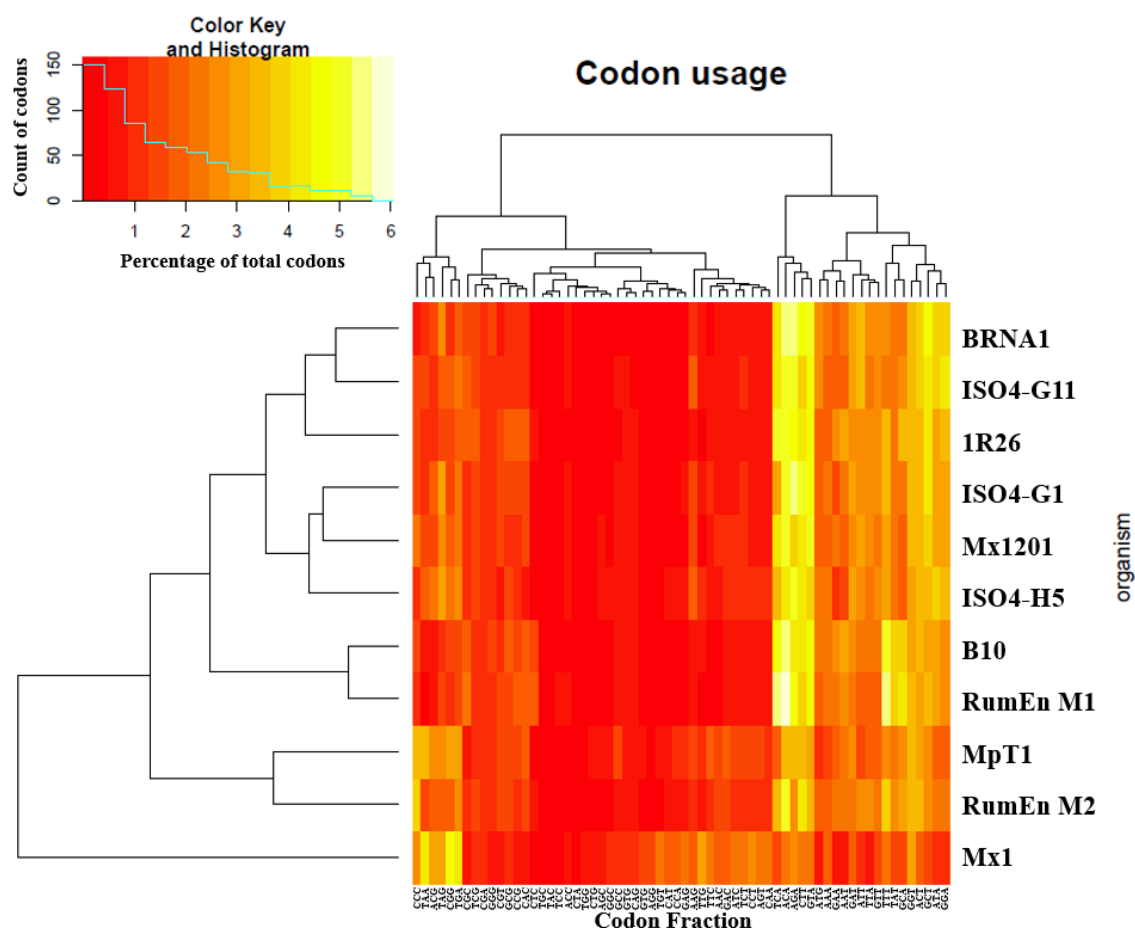
Ten Methanomassiliicoccales genomes analysed are predicted to contain a tRNA gene for all 21 amino acids (Table 4.3), the amber suppressor tRNA<sup>Pyl</sup> is not predicted in the RumEn M2 genome. Seven genomes are predicted to contain 47 tRNAs, four genomes are predicted to contain 45 to 49 tRNAs, this is due to the variation in copy numbers of tRNA<sup>Leu</sup>, tRNA<sup>Lys</sup>, tRNA<sup>Pro</sup> and tRNA<sup>Thr</sup> in the Mx1, RumEn M1, RumEn M2 and 1R26 genome. The numbers of tRNAs corresponding to each amino acid is highly conserved between the Methanomassiliicoccales genomes, with some tRNAs in high numbers and some tRNA genes in low numbers, as shown in Table 4.3. Introns have been found consistently within tRNA<sup>Trp</sup> and tRNA<sup>Tyr</sup>.

**Table 4.3 Predicted tRNAs**

tRNA corresponding Amino acid	ISO4- H5	ISO4- G1	ISO4- G11	BRNA1	RumEn M1	RumEn M2	1R26	Mx1201	Mx1	B10	MpT1
Alanine	3	3	3	3	3	3	3	3	3	3	3
Arginine	5	5	5	5	5**	5**	5	5	6**	4*	5
Asparagine	1	1	1	1	1*	1*	1	1	1*	1*	1
Aspartate	1	1	1	1	1	1	1	1	1	2	1
Cysteine	1	1	1	1	1*	1*	1	1	1*	1*	1
Glutamine	2	2	2	2	2	2	2	2	2	2	2
Glutamate	2	2	2	2	2	2	2	2	2	3*	2
Glycine	3	3	3	3	3	3	3	3	3	3	3
Histidine	1	1	1	1	1	1	1	1	1	1	1
Isoleucine	1	1	1	1	1	1	1	1	1	2	1
Leucine	5	5	5	5	5	5	6	5*	4	4	5*
Lysine	2	2	2	2	1	1	2	2	1	1	2
Methionine	3*	3*	3*	3*	3**	3**	3*	3*	3	2	3*
Phenylalanine	1	1	1	1	1	1	1	1	1	1	1
Proline	3	3	3	3	3	3	2	3	3	3	3
Pyrrolysine	1	1	1	1	1	0	1	1	1	1	1
Serine	4	4	4	4	4*	4*	4	4	4*	4*	4
Threonine	3	3	3	3	3	3	4	3	3	3	3
Tryptophan	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*
Tyrosine	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*
Valine	3	3	3	3	3	3	3	3	3	3	3
Pseudo	1	0	0	0	0	0	1	0	0	1*	0
Total	47	47	47	47	46	45	49	47	46	47	47
Intron-containing tRNAs	3	3	3	3	8	9	3	4	7	8	4

\*intron containing tRNAs

*Codon and amino acid usage.* The codon usage for each genome is summarised in Table A.4.1, and visualized in Figure 4.2. The overall codon usage pattern is similar between ISO4-G1, Mx1201, ISO4-H5, ISO4-G11, BRNA1, 1R26, RumEn M1, RumEn M2 and B10 (Figure 4.2), while Mx1 and MpT1 display different patterns. Mx1 codon usage is likely a result of its low G+C content of 41.3% as reflected in the codon usage for the amino acids phenylalanine, asparagine, lysine, glutamate, tyrosine, alanine and cysteine, where Mx1 preferentially uses codons with a lower G or C content (Table A.4.1). RumEn M1 has the highest G+C content (62.1%) of the genomes analysed, and GGC and GGG are preferred over GGA and GGU to encode glycine, GUG and GUC are preferred over GUU and GUA to encode valine (Table A.4.1). No correlation of G+C content and codon preference were observed in MpT1 (Table A.4.1).



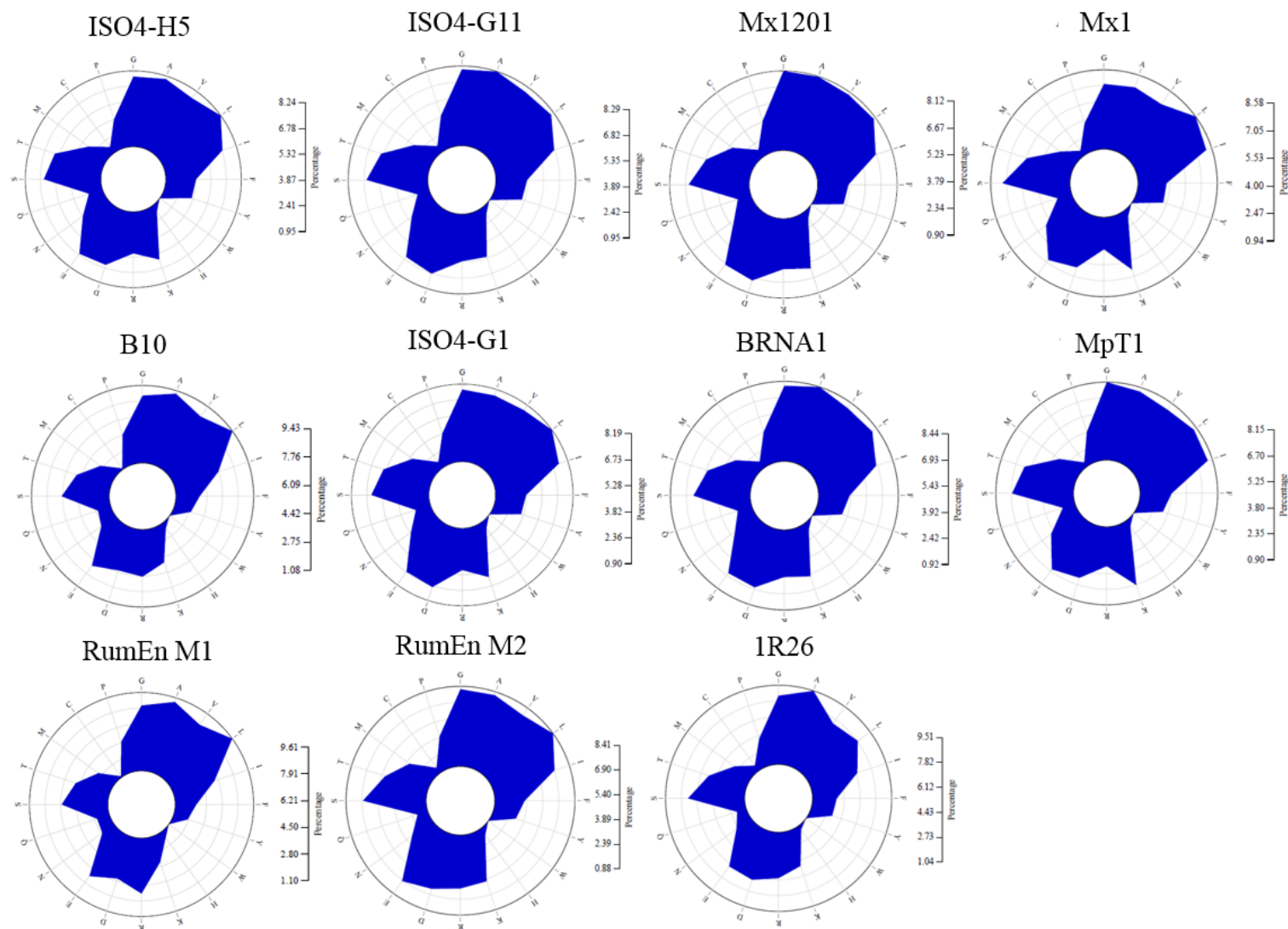
**Figure 4.2 Codon usage heatmap of the Methanomassiliicoccales genomes analysed.** The color key on the top left hand corner, indicates codons with the lowest percentage codons in red, and most abundant codons are displayed in pale yellow, while the histogram displays the count of codons in 0.37% intervals across genomes.

Translational start and stop codons appear to differ among the genomes investigated. B10 has the lowest usage of methionine as a translational initiator, and the highest usage of valine as a translational initiator, whereas BRNA1 has the highest usage of leucine as a translational initiator (Table A.4.1). All of the genomes analysed have a high preference for using the opal codon as a stop codon, with the exception of Mx1 genome which uses the opal codon and ochre codon at similar levels (Table A.4.1). The amber codon is associated with pyrrolysine incorporation in all genomes examined.

Despite the differences in codon usage, the genomes investigated are rather conserved in their amino acid usage pattern (Figure 4.3). All 11 genomes analysed have a high percentage of branched-chain amino acids, glycine, alanine and serine, and uses 4% to 7% of charged amino acids glutamate, lysine, arginine and aspartate. The least used amino acid is tryptophan for all the genomes analysed (Figure 4.3). The usage of cysteine and histidine is also quite low in all

11 genomes. These three low frequency amino acids only have one single tRNA associated with each of them.

*CRISPR elements.* The CRISPR elements of 11 Methanomassiliicoccales genomes were predicted and summarised in Table 4.1, the CRISPR associated genes were summarised in Table A.4.2. Based on the CRISPR associated genes, the ISO4-H5 genome may possess a type I-E CRISPR element, 8 other genomes were predicted with *cas3* gene homologue, which suggests they may also be type I CRISPR element, however, the subtypes cannot be determined due to lack of other associated gene homologues.



**Figure 4.3 Amino acid usage map of the *Methanomassiliicoccales* genomes analysed.** Pyrrolysine is not included in this figure. The amino acid usage is displayed as percentages, scales are different for every genome and are displayed to the right of each diagram.



### 4.2.3 Pyrrolysine usage

All of the genes required for pyrrolysine biosynthesis and incorporation are found among ten of the Methanomassiliicoccales genomes and the structure of the pyrrolysine operon is conserved (Figure 4.4), only the incomplete RumEn M2 genome lacks the pyrrolysine biosynthesis genes. The B10 genome contains two pyrrolysine operons located in two different contigs. The pyrrolysine operon is located adjacent to *mtmB*, *mtbB*, *mttB* and other genes involved in methylamine utilisation, including corrinoid proteins associated with each methyltransferase *mtmC*, *mtbC*, *mttC*, methylamine permeases *mtbP* and *mttP* and methyltransferase corrinoid activation protein, *ramA* (Figure 4.4). The genes containing in-frame amber codons were examined to predict pyrrolysine incorporation in ISO4-H5, ISO4-G11, ISO4-G1, BRNA1, RumEn M1, RumEn M2, 1R26 and MpT1, and were classified into three categories: Class 1 genes which have amber codon read-through and predicted incorporation of pyrrolysine; Class 2 genes that were predicted to use the amber codon as a stop codon; Class 3 genes with uncertain amber codon usage due to lack of homologous genes. The amber codon usage and putative Pyl-containing proteins in the Methanomassiliicoccales genomes analysed are listed in Table 4.4 (also see Tables A.4.1, A.4.3 – A.4.9) and are displayed in Figure 4.4.

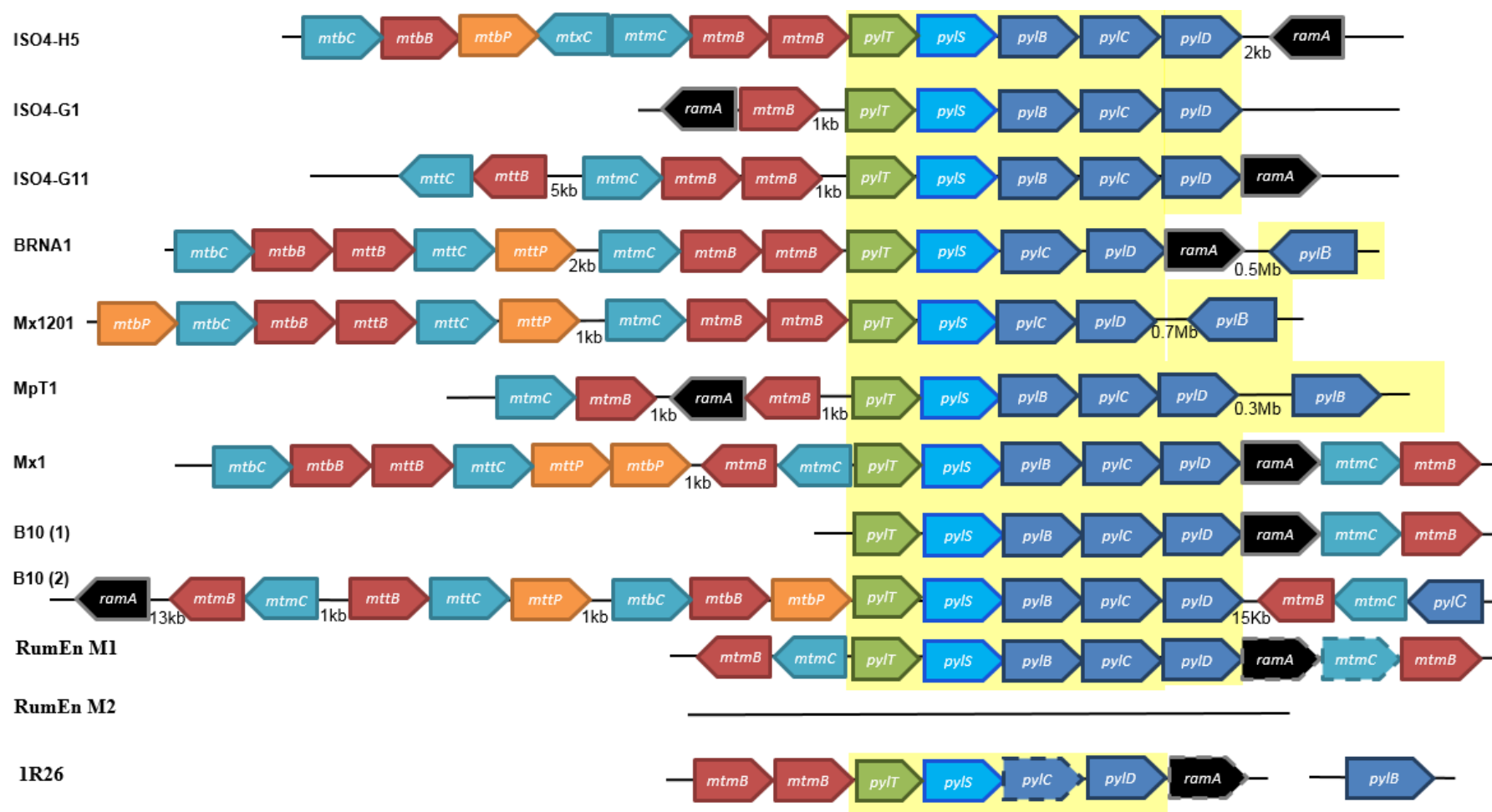
The Methanomassiliicoccales genomes fall into two different patterns of pyrrolysine utilisation. The genomes of B10, Mx1, RumEn M1, RumEn M2 and MpT1 are predicted to contain very few proteins incorporating pyrrolysine outside of MtmB, MtbB and MttB, whereas ISO4-H5, ISO4-G1, ISO4-G11, Mx1201, 1R26 and BRNA1 are predicted to incorporate pyrrolysine at a relatively higher frequency (Table 4.4, Figure 4.5). The genomes of B10 and Mx1 both have a high number of genes containing an in-frame amber codon, however other than the MtmB/MtbB/MttB genes, only a putative Fe-S binding protein (AGY50215), is predicted to incorporate pyrrolysine, with the remainder predicted to use the amber codon as a translational stop. The remaining Methanomassiliicoccales genomes have a relatively high number of genes predicted to use pyrrolysine and a low number of genes predicted to contain an in-frame amber codon.

**Table 4.4 Amber codon utilisation in Methanomassiliicoccales**

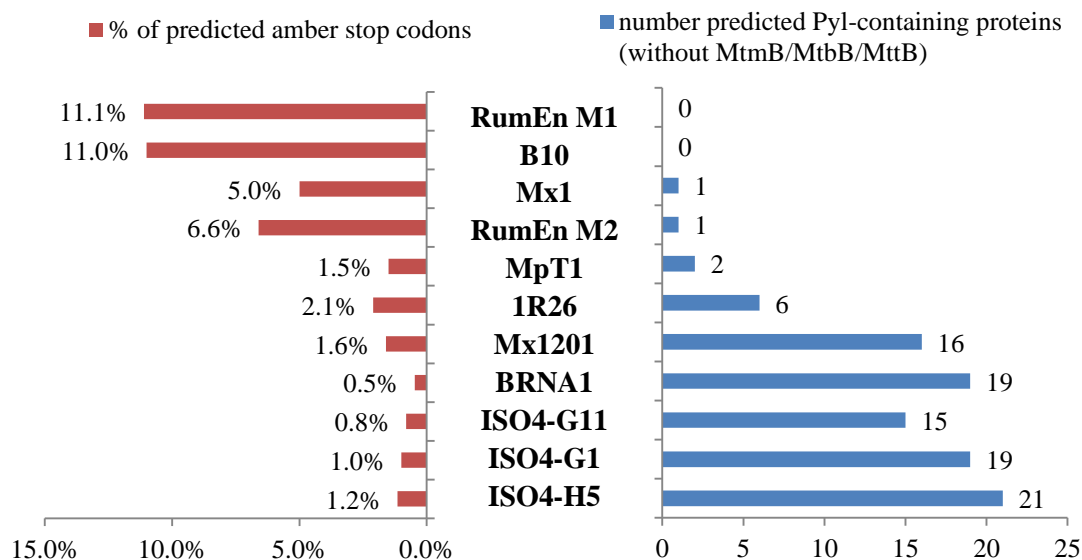
Genomes	CDS Class 1 (%)*	CDS Class 2 (%)	CDS Class 3 (%)	Total
ISO4-H5	25 (1.4)	10 (0.6)	11 (0.6)	46 (2.5)
ISO4-G1	28 (1.9)	6 (0.4)	9 (0.6)	43 (2.9)
ISO4-G11	20 (1.1)	5 (0.3)	9 (0.5)	34 (2.0)
BRNA1	25 (1.7)	6 (0.4)	1 (0.1)	32 (2.1)
RumEn M1	2 (0.1)	252 (11.1)	41 (1.8)	295 (13.0)
RumEn M2	1 (0.1)	90 (6.6)	22 (1.6)	113 (8.2)
IR26	10 (0.6)	37 (2.1)	27 (1.5)	74 (4.1)
Mx1201	20 (1.2)	26 (1.6)	Nd	46 (2.8)
Mx1	5 (0.3)	92 (5.0)	Nd	97 (5.3)
B10	7 (0.3)	293 (11.0)	Nd	300 (11.3)
MpT1	4 (0.3)	13 (0.9)	8 (0.6)	25 (1.8)

\*Number of CDS after read-through is merged, including pseudogenes. Nd: not determined in this study.

MpT1 has a very low number of genes containing in-frame amber codons. Only two proteins encoded by these genes, other than the MtmB/MtbB/MttB, are predicted to incorporate pyrrolysine (Table A.4.5). Additionally, the amber-codon containing genes are either hypothetical proteins, mobile elements or genes with a very short distance between the amber codon and the next stop codon.



**Figure 4.4 Gene organisation of the pyrrolysine operon in *Methanomassiliicoccales* genomes and their co-localisation with methylamine utilisation genes.** The pyrrolysine operon is shaded in yellow and includes: Amber suppressor tRNA<sup>CUA</sup> (*pylT*). Pyrrolysine-tRNA ligase (*pylS*). Methylornithine synthase, *pylB*; (2*R*,3*R*)-3-methylornithyl-*N*<sup>6</sup>-lysine synthase (*pylC*) and pyrrolysine synthase (*pylD*). Methylamine utilisation genes include: Monomethylamine:corrinoide methyltransferase (*mtmB*), dimethylamine:corrinoide methyltransferase (*mtbB*), trimethylamine:corrinoide methyltransferase (*mttB*), monomethylamine methyltransferase corrinoide protein (*mtmC*), dimethylamine methyltransferase corrinoide protein (*mtbC*), trimethylamine methyltransferase corrinoide protein (*mttC*), dimethylamine permease (*mtbP*), trimethylamine permease (*mttP*), methyltransferase corrinoide activation protein (*ramA*). Genes with dashed border were pseudogenes. Gene sizes are not drawn to scale.



**Figure 4.5 Comparison of predicted pyrrolysine containing proteins (excluding MtmB/MtbB/MttB) to the percentage of predicted amber stop codons.**

#### 4.2.4 Secretome

The predicted secretome (surface associated and/or released from the cell) proteins of the 11 genomes are summarised in Table 4.5 and Table A.4.6. The genomes analysed are predicted to devote between 2.5% and 5.6% of their ORFs to the export of extracellular and surface-associated proteins. A large proportion of the secretomes examined are predicted to have ORFs with motifs or domains for attachment to the cell surface. A range of between 16 to 75 genes per genome are predicted to be membrane anchored by one to two transmembrane helices or by an N-terminal lipobox. Around 4 to 25 genes per genome are predicted to be either unattached (i.e. secreted) or associated with the cell wall using other (unknown) mechanisms. The extracellular proteins were analysed for the presence of repeat domains and periplasmic domains (the periplasmic domain belongs to the substrate binding subunit of ABC transporters) and the results are listed in Table A.4.10.

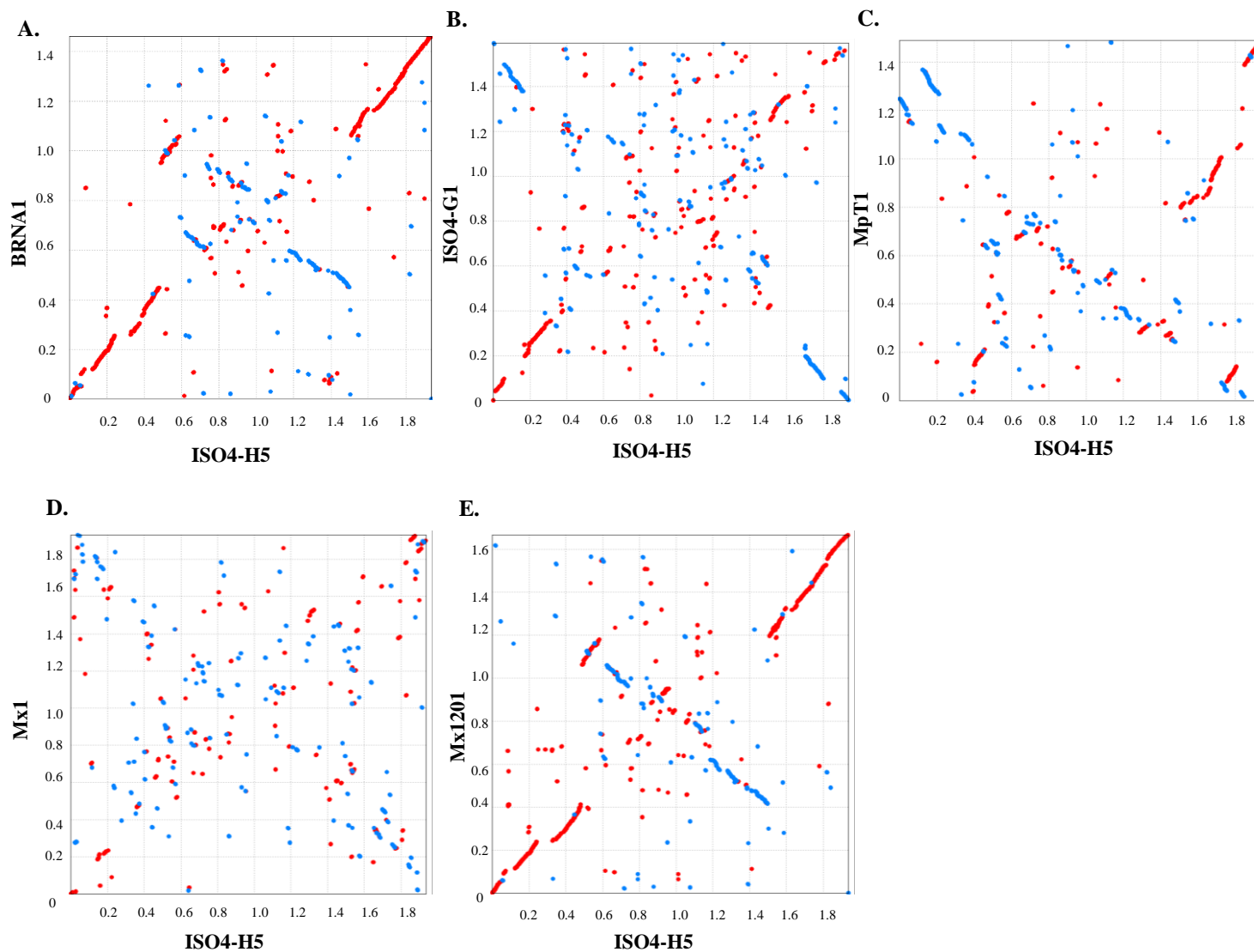
**Table 4.5 Predicted secretome of Methanomassiliicoccales**

Secretome	ISO4-H5	ISO4-G1	ISO4-G11	BRNA1	RumEn M1	RumEn M2	1R26	Mx1201	Mx1	B10	MpT1
Integral membrane protein	13	12	15	15	18	8	13	13	15	25	9
Lipobox	6	2	8	3	9	2	13	2	4	0	1
Two TMH	22	14	17	12	12	6	10	10	31	23	29
Membrane anchor	11	6	7	4	3	1	3	6	3	13	2
C-terminal	35	27	36	21	24	7	28	10	35	39	17
membrane anchor	15	8	9	5	25	10	15	13	9	24	4
Secreted	102	69	92	60	91	34	82	54	97	124	62
Total	(5.6%)	(4.5%)	(5.2%)	(3.8%)	(4.0%)	(2.5%)	(4.6%)	(3.2%)	(5.2%)	(4.6%)	(4.4%)
(% of total ORFs)											

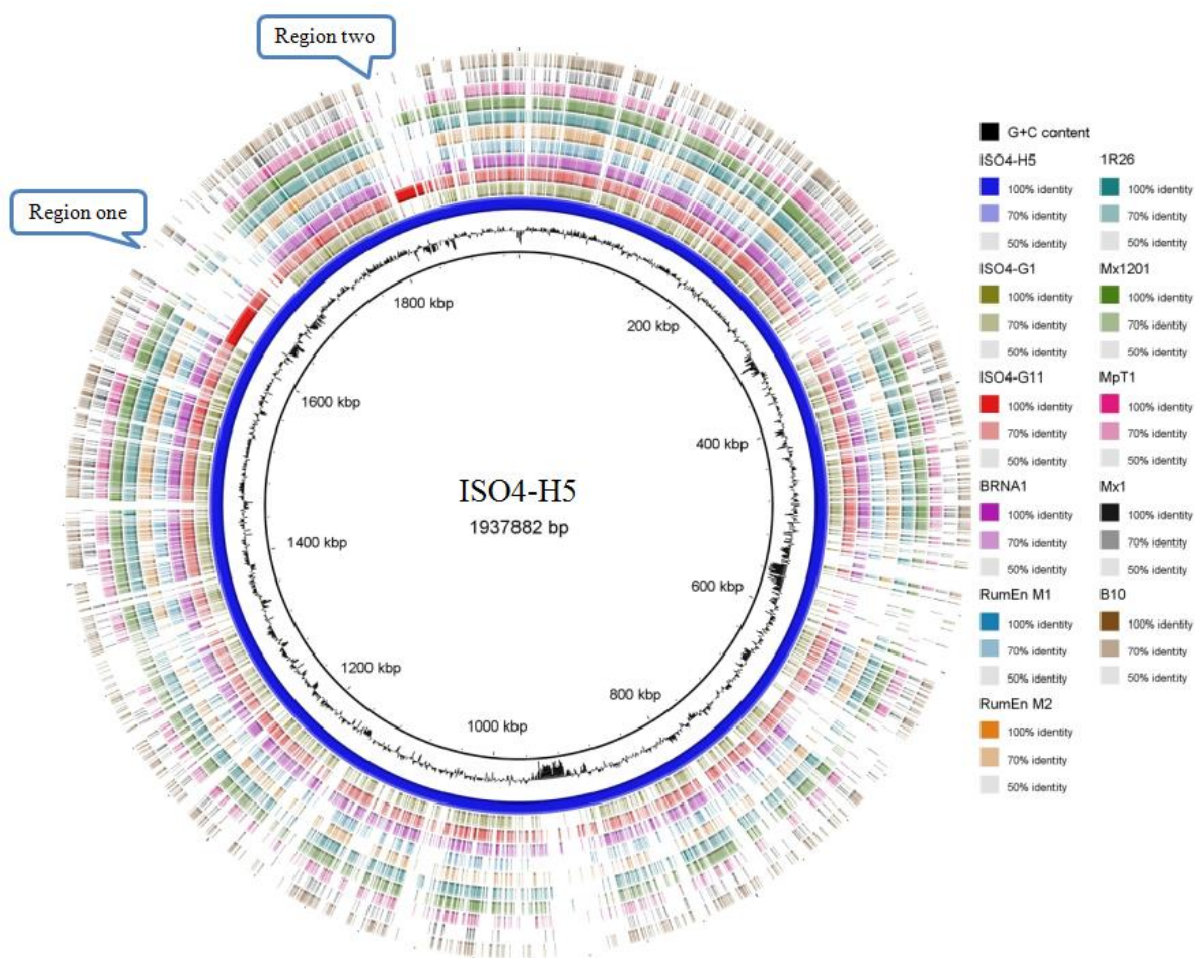
Prediction E value cutoff  $1 \times 10^{-05}$ .

#### 4.2.5 Synteny

ISO4-H5 shares a high degree of synteny and displays an X-shaped alignment with BRNA1 and Mx1201 (Figure 4.6A, 4.6E). A comparison of the nucleotide identity across all sequenced genomes using ISO4-H5 as the reference (Figure 4.7, Figure A.4.3F) revealed two regions of 100% identity to the genome of ISO4-G11. Region 1 encompasses 30.3 kbp and Region 2 encompasses 26.3 kbp in the ISO4-H5 genome, and the genes in both regions are summarised in Table 4.6



**Figure 4.6 Gene synteny plots for completed genomes of Methanomassiliicoccales.** PROmer alignments of the ISO4-H5 genome against genomes of other Methanomassiliicoccales are shown. The alignments were plotted using MUMmer (Delcher *et al.* 2003) with forward matches shown in red and reverse matches in blue. The units displayed on both axes are in Mb.



**Figure 4.7 Genome map comparison of members of the order Methanomassiliicoccales.** Methanomassiliicoccales genomes are plotted using ISO4-H5 as a reference. The colors in each genome ring accord with the identity illustrated in the legend. Circles are ordered 1 (innermost) to 10 (outermost). Circle 1: DNA base pairs of ISO4-H5. Circle 2: G+C content of ISO4-H5 genome. Circle 3: ISO4-H5 genome nucleotide sequence identity to itself. Circles 4 to 10: ISO4-H5 genome identity to ISO4-G1, ISO4-G11, BRNA1, RumEn M1, RumEn M2, 1R26, Mx1201, MpT1, Mx1, and B10, genomes respectively.

**Table 4.6 Gene content of Regions 1 and 2 in ISO4-H5 and ISO4-G11 genomes**

ISO4-H5	ISO4-G11	Predicted gene product
<b>Region 1</b>		
AR505_1521	ISO4G11_1250	Adhesin-like protein
AR505_1522	ISO4G11_1251	Hypothetical transmembrane protein
AR505_1523	ISO4G11_1252	ABC transporter substrate binding protein
AR505_1524	ISO4G11_1253	Adhesin-like protein
AR505_1525	ISO4G11_1254	ABC transporter permease protein
AR505_1526	ISO4G11_1255	ABC transporter ATP-binding protein
AR505_1527	ISO4G11_1256	Hypothetical protein
AR505_1528	ISO4G11_1257	Hypothetical protein
AR505_1529	ISO4G11_1258	Hypothetical protein
AR505_1530	ISO4G11_1259	Adhesin-like protein
AR505_1531	ISO4G11_1260	ATP-dependent DNA helicase
AR505_1532	ISO4G11_1261	Hypothetical transmembrane protein
AR505_1533	ISO4G11_1262	Adhesin-like protein
AR505_1534	ISO4G11_1263	Adhesin-like protein
AR505_1535	ISO4G11_1264	Hypothetical protein
AR505_1536	ISO4G11_1266	Hypothetical protein
AR505_1537	ISO4G11_1267	Hypothetical protein
<b>Region 2</b>		
AR505_1705	ISO4G11_1645	ABC transporter ATP-binding protein
AR505_1706	ISO4G11_1644	ABC transporter permease protein
AR505_1707	ISO4G11_1643	Adhesin-like protein
AR505_1708	ISO4G11_1642	ABC transporter substrate binding protein
AR505_1709	ISO4G11_1641	Hypothetical transmembrane protein
AR505_1710	ISO4G11_1640	Hypothetical transmembrane protein
AR505_1711	ISO4G11_1639	Adhesin-like protein
AR505_1712	ISO4G11_1638	Adhesin-like protein
AR505_1713	ISO4G11_1637	Adhesin-like protein
AR505_1714	ISO4G11_1636	Hypothetical protein
AR505_1715	ISO4G11_1635	Hypothetical protein
AR505_1716	ISO4G11_1634	Hypothetical protein
AR505_1717	ISO4G11_1633	Hypothetical protein
AR505_1718	ISO4G11_1632	Hypothetical protein

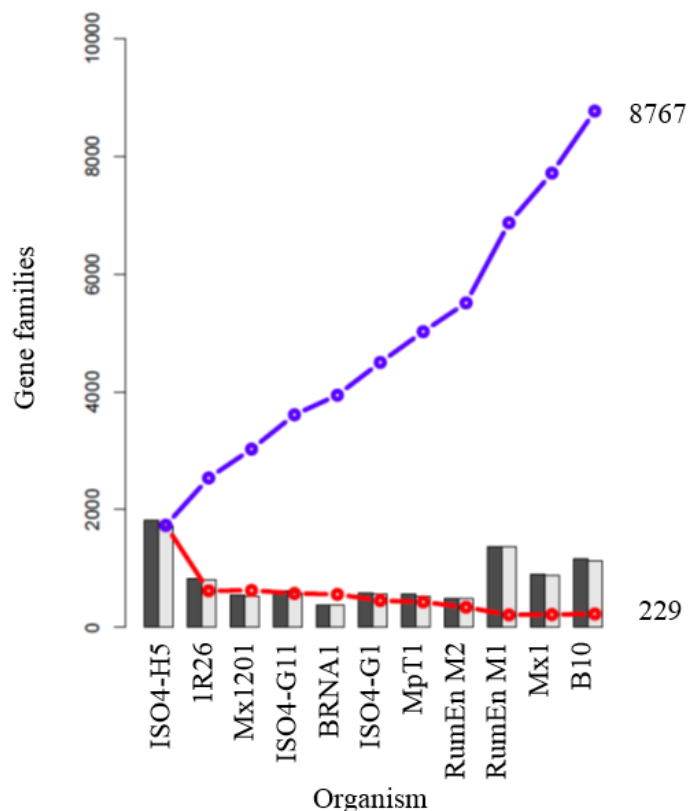
In order to identify similar regions between other members of Methanomassiliicoccales, each genome was used as reference genome and the nucleotide similarity was displayed (Figure A.4.3). A less conserved, smaller genomic island of approximately 5 kb was detected between BRNA1 and ISO4-G11, however, the alignment ended prematurely as it was located at the end of a contig in the ISO4-G11 draft genome (Figure A.4.4). The two conserved genes found in this region are a Type I site-specific deoxyribonuclease and a hypothetical protein.

#### 4.2.6 Comparative analysis of gene families

The term core-genome was used to describe the genes or gene families present in all completed genomes analysed while the order level pan-genome was used to define the full complement of genes or gene families present in all sequenced genomes. When only completed genomes are analysed, the core genome consists of 415 gene families (Figure 4.10A). When all genomes are analysed, 229 gene families are conserved between all genomes analysed (Figure 4.10C) and the order level pan-genome consists of 8767 gene families (Figure 4.8). The inclusion of draft genomes into analysis revealed 78 gene families conserved between members of GIT1, and 226 gene families conserved between members of GIT2 (Figure A.4.5), during this analysis

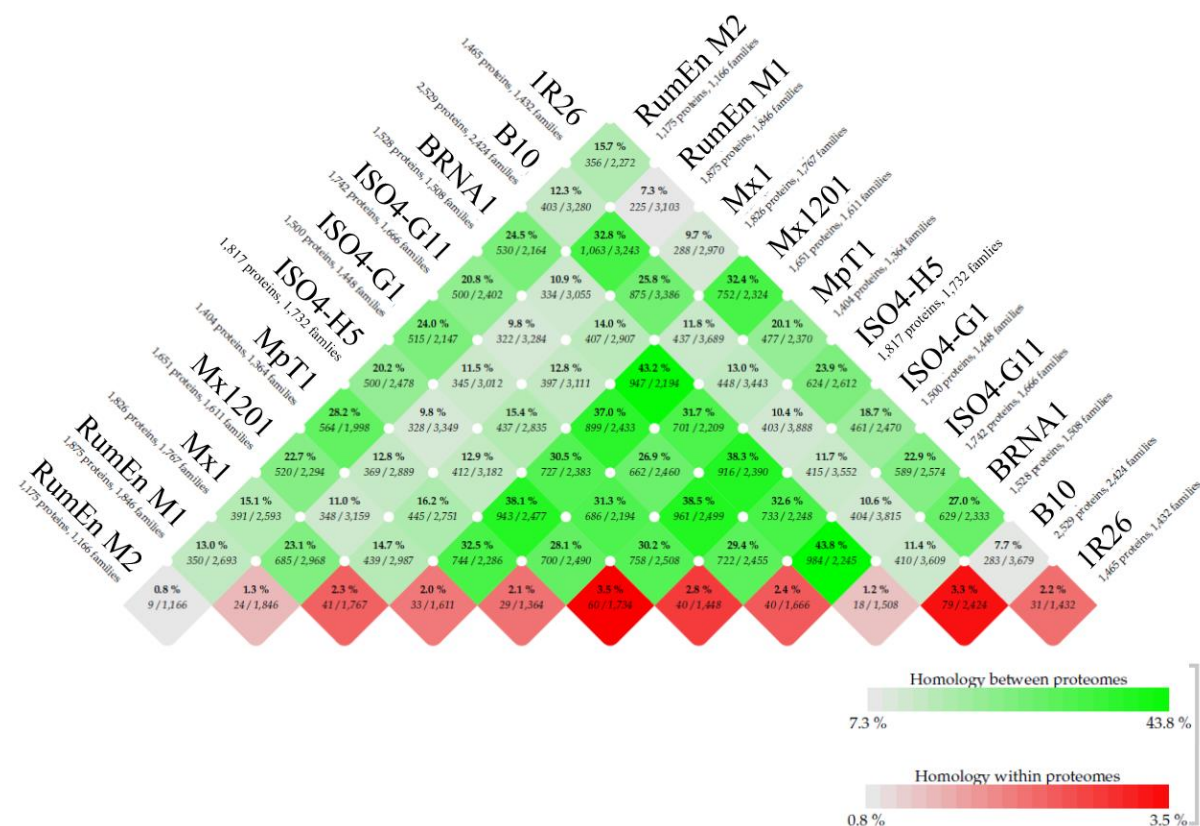


it was identified that the incomplete genomes may confound the gene conservation analysis due to low level completion and poor quality sequence, the RumEn M1, RumEn M2 and 1R26 genomes are only 96.6%, 94.2% and 93% complete respectively, with 396, 194 and 315 predicted pseudogenes respectively, including some functionally important and conserved genes, for example the *mrtA* gene was marked as a pseudogene in the RumEn M1 genome, the *mrtG* gene was entirely absent from the 1R26 annotation due to a lack of a stop codon. To prevent false negatives introduced by the incomplete genomes, only the completed genomes were analysed to to identify the core genome and environmental conserved gene families.



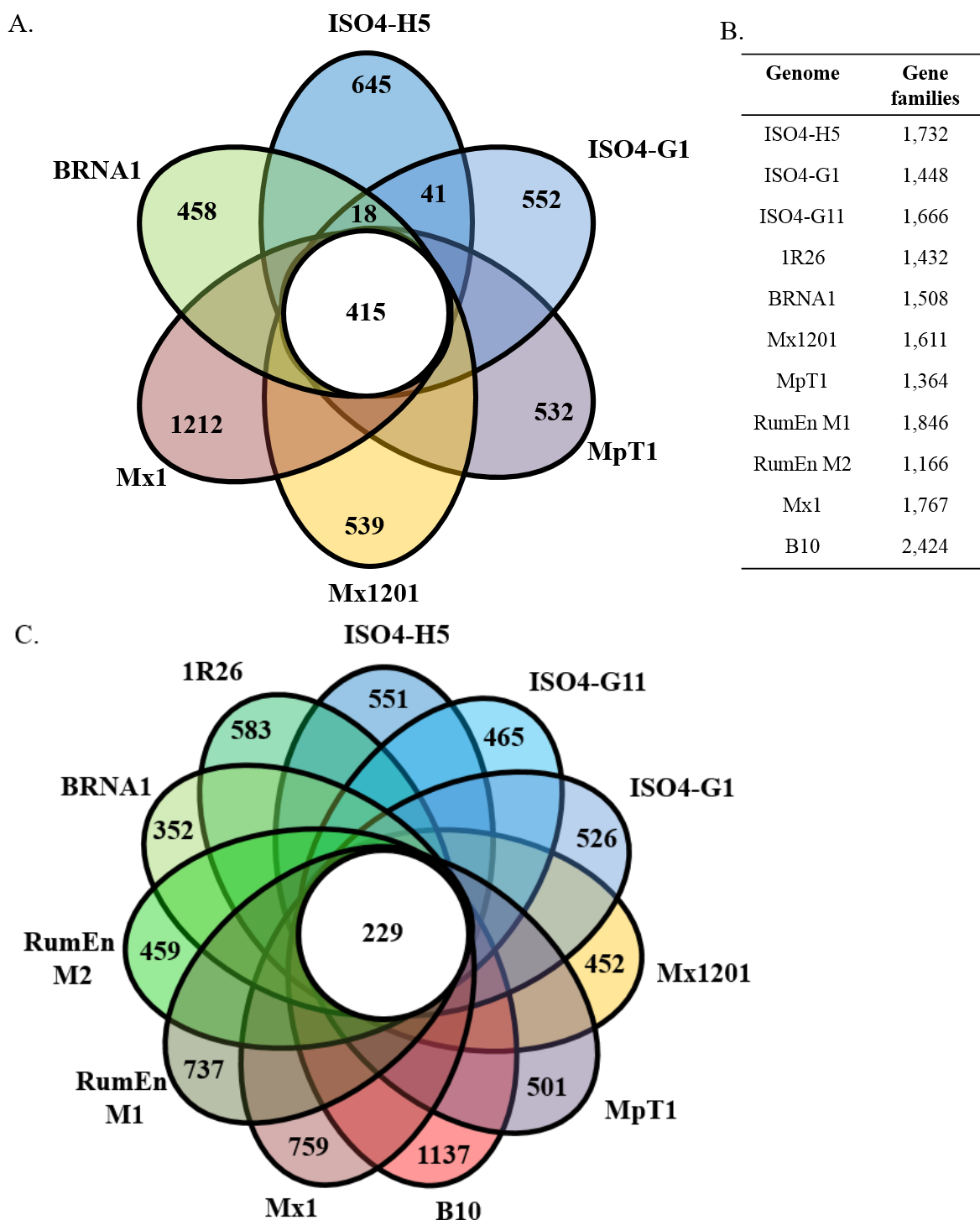
**Figure 4.8 Order level pan-genome and core-genome plot of Methanomassiliicoccales genomes.** The blue line indicates the cumulative order level pan-genome, and the red line indicates the core-genome. Dark grey bars indicate the number of new genes with the addition of each genome across the x-axis; the light grey bar indicates new gene families.

The pattern of gene family conservation as illustrated in Figure 4.9 show that six members of the Methanomassiliicoccales are more similar to each other, while B10, RumEn M1 and Mx1 are more similar to each other. The analysis shows that ISO4-H5 has the highest percentage of paralogs. The highest predicted gene family conservation is between ISO4-G11 and BRNA1 at 44%.



**Figure 4.9 BLASTP matrix illustrating the number of conserved protein families between Methanomassiliicoccales genomes.** Conservation is defined as 50% coverage and 50% identity. The color intensities are based on the relative percentage of conserved gene families; green depicts conserved protein families between genomes, red depicts protein homology within genome.

The number of gene families in the core-genome corresponds with the phylogenetic relationship between the organisms, as does the number of gene families specific to each genome. The number of gene families specific to RumEn M1, Mx1 and B10 are high, with 40%, 43% and 46.9% of their total gene families being specific to their genomes, respectively (Figure 4.10B).



**Figure 4.10 Conserved and novel gene families among the 11 Methanomassiliicoccales genomes.** **A.** Venn diagram indicating the number of conserved gene families between completed genomes based on BLASTp analysis, using a 50% identity and 50% coverage cutoff. Regions that do not overlap with other genomes depict the number of unique gene families under the same criteria. **B.** Table listing the total number of gene families in each of the Methanomassiliicoccales genomes. **C.** Venn diagram indicating the number of conserved gene families between all genomes analysed based on BLASTp analysis, using a 50% identity and 50% coverage cutoff. Regions that do not overlap with other genomes depict the number of unique gene families under the same criteria.

*Analysis of the Methanomassiliicoccales core genome.* The number of gene families in the core genome is summarised in Figure 4.11 and Table 4.7. A full list of conserved gene families is presented in Appendix (Table A.4.11). The core genome includes an operon of type III ribulose biphosphate carboxylase (*rbcL*), ribose-1,5-biphosphate isomerase (*e2b2*) and AMP phosphorylase (*deoA*) genes.

**Table 4.7 ISO4-H5 conserved gene families and novel gene families classified by COG functional categories**

COG	Code	Gene families				
		Total	Core	Rumen	Ovine	Unique
Translation	J	144	113	0	0	3
RNA processing and modification	A	1	0	0	0	0
Transcription	K	68	21	3	0	13
Replication, recombination and repair	L	123	17	1	0	38
Chromatin structure and dynamics	B	1	0	0	0	0
Cell cycle control, mitosis and meiosis	D	9	2	0	0	2
Nuclear structure	Y	0	0	0	0	0
Defense mechanisms	V	17	3	0	0	2
Signal transduction mechanisms	T	20	3	1	0	9
Cell wall/membrane biogenesis	M	27	2	0	0	11
Cell motility	N	2	0	0	0	2
Cytoskeleton	Z	0	0	0	0	0
Extracellular structures	W	0	0	0	0	0
Intracellular trafficking and secretion	U	13	6	0	0	0
Post-translational modification, protein turnover, chaperones	O	51	20	0	0	12
Energy production and conversion	C	89	34	1	4	4
Carbohydrate transport and metabolism	G	40	15	2	0	4
Amino acid transport and metabolism	E	99	34	0	2	7
Nucleotide transport and metabolism	F	46	29	1	2	1
Coenzyme transport and metabolism	H	104	31	0	4	8
Lipid transport and metabolism	I	18	9	1	1	6
Inorganic ion transport and metabolism	P	89	2	1	7	14
Secondary metabolites biosynthesis, transport and catabolism	Q	14	0	0	0	2
General function prediction only	R	241	37	1	4	73
Function unknown	S	126	24	0	3	24
Unclassified COG	Unclassified COG	91	6	2	3	40
Not in COGs	-	393	3	4	11	276

**Figure 10: Distribution of COG categories for the 1000 most abundant proteins in *E. coli*.**

COG Category	Percentage
[K] Transcription	27%
[R] General function prediction only	10%
[C] Energy production and conversion	9%
[L] Replication, recombination and repair	8%
[U] Intracellular trafficking, secretion, and vesicular transport	7%
[I] Lipid transport and metabolism	7%
[F] Nucleotide transport and metabolism	7%
[T] Signal transduction mechanisms	6%
[V] Defense mechanisms	6%
[O] Post-translational modification, protein turnover, and chaperones	5%
[S] Function unknown	5%
[G] Carbohydrate transport and metabolism	4%
[E] Amino acid transport and metabolism	4%
[H] Coenzyme transport and metabolism	2%
[D] Cell cycle control, cell division, chromosome partitioning	2%
[M] Cell wall/membrane/envelope biogenesis	1%
[P] Inorganic ion transport and metabolism	1%
[N] Not in COGs	1%
[A] Unclassified COG	1%
[B] Cell motility	1%
[J] Translation	0%

*Adaptation to the rumen environment.* 18 gene families are uniquely conserved between all four genomes of Methanomassiliicoccales sourced from the rumen (Table 4.8) while 41 gene families are conserved between the ISO4-H5 and ISO4-G1 genomes, both of which have been sequenced from the ovine rumen (Table 4.9).

**Table 4.8 Gene families conserved only in rumen by functional category**

Locus_tag	Predicted gene product	COG category
AR505_0479	hypothetical protein	Not in COGs
AR505_0645	carbon starvation protein CstA	[T]
AR505_0681	4Fe-4S ferredoxin iron-sulfur binding domain-containing protein	[C]
AR505_0718	TfoX N-terminal domain protein	[K]
AR505_0801	HTH domain-containing protein	[K]
AR505_0806	MFS transporter	[P]
AR505_0831	transcriptional regulator MarR family	[K]
AR505_0847	hypothetical protein	Not in COGs
AR505_0921	hypothetical protein	[unclassified COG]
AR505_0960	Arylsulfotransferase AssT	[unclassified COG]
AR505_0995	xanthine/uracil permease family protein	[R]
AR505_1037	DNA polymerase IV	[L]
AR505_1083	hypothetical protein	Not in COGs
AR505_1104	phosphoenolpyruvate synthase PpsA2	[G]
AR505_1105	hypothetical protein	Not in COGs
AR505_1106	dolichol kinase (7 TMHs)	[I]
AR505_1118	MTA/SAH nucleosidase MtnN	[F]
AR505_1272	transporter MIP family	[G]

\*only one gene from ISO4-H5 for each gene family is represented

**Table 4.9 Gene families conserved only in ovine rumen by functional category**

Locus_tag	Predicted gene product	COG category
AR505_0100	thiamine monophosphate synthase ThiE1	[H]
AR505_0103	thiamine biosynthesis protein ThiF1	[H]
AR505_0104	thiamine biosynthesis protein ThiS	[H]
AR505_0112	ABC transporter substrate-binding protein	[P]
AR505_0314	hypothetical protein	Not in COGs
AR505_0320	hypothetical protein	Not in COGs
AR505_0322	hypothetical protein	Not in COGs
AR505_0323	hypothetical protein	Not in COGs
AR505_0346	hypothetical protein	Not in COGs
AR505_0597	anaerobic cobalt chelatase CbiK	[H]
AR505_0629	transmembrane protein	Not in COGs
AR505_0678	Citrate synthase GltA	[C]
AR505_0707	hypothetical protein	Not in COGs
AR505_0720	hypothetical protein	[S]
AR505_0725	4Fe-4S ferredoxin iron-sulfur binding domain-containing protein	[C]
AR505_0726	ApbE family protein	[S]
AR505_0812	Na/Pi-cotransporter II-like protein	[P]
AR505_0822	Nitroreductase family protein	[C]
AR505_0835	transporter Na <sup>+</sup> /H <sup>+</sup> antiporter family	[P]
AR505_0875	hypothetical protein	Not in COGs
AR505_0933	Cardiolipin synthase	[I]
AR505_0941	MFS transporter	[P]
AR505_0978	flavodoxin-like protein	[C]
AR505_0983	ATPase	[R]
AR505_1039	NADPH-dependent FMN reductase	[R]
AR505_1058	hypothetical protein	Not in COGs
AR505_1081	small multidrug resistance protein	[P]
AR505_1082	small multidrug resistance protein	[P]
AR505_1088	hypothetical protein	Not in COGs
AR505_1100	K <sup>+</sup> -dependent Na <sup>+</sup> /Ca <sup>2+</sup> exchanger	[P]
AR505_1102	thymidylate kinase Tmk1	[F]
AR505_1114	hypothetical protein	[unclassified COG]
AR505_1115	DEAD/DEAH box helicase domain-containing protein	[R]
AR505_1178	radical SAM domain containing protein	[R]
AR505_1327	monomethylamine methyltransferase MtmB1	[unclassified COG]
AR505_1363	hypothetical protein	[unclassified COG]
AR505_1464	L-2,4-diaminobutyrate decarboxylase	[E]
AR505_1465	2,4-diaminobutyrate 4-transaminase	[E]
AR505_1470	transmembrane protein	Not in COGs
AR505_1740	Fic family protein	[S]
AR505_1743	5'-nucleotidase SurE	[F]

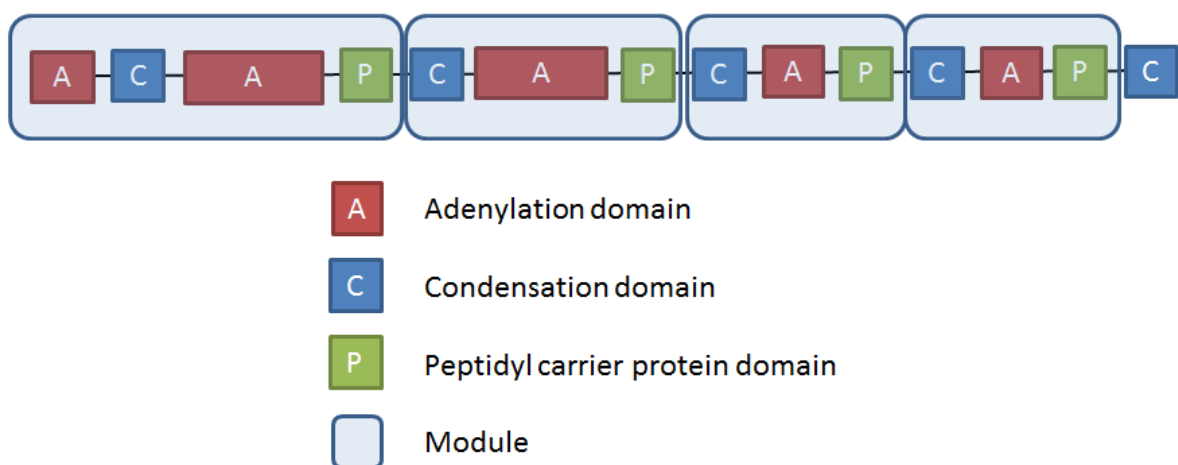
\*only one gene from each gene family is represented

*Novel gene families in genomes of ovine rumen strains.* The analysis of gene family conservation identified 551 unique gene families in ISO4-H5 (Table A.4.12), 526 in ISO4-G1 (Table A.4.13) and 465 in ISO4-G11 (Table A.4.14). Among the 551 gene families novel to ISO4-H5, 343 are predicted to be hypothetical proteins. ISO4-H5 contains a high number of insertion elements (Table 3.6), consisting of nine gene families out of the 40 gene families categorised under the replication, recombination and repair COG function. ISO4-H5 carries two copies of the bifunctional dihydrofolate synthetase *folC* gene involved in tetrahydrofolate production. One copy (AR505\_0729) is considered novel under the criteria used in this study. ISO4-H5 carries two copies of the cell division protein *ftsZ* gene involved in cell replication, one (AR505\_1792) is considered novel under the criteria used in this study. Three copies of signal peptidase I have been found in ISO4-H5 genome, and one (AR505\_1640) is considered to be novel under the criteria of this study.

Among the 465 gene families novel to ISO4-G11, 292 are predicted to be hypothetical proteins, and 73 gene families are poorly characterized. There are 130 gene families assigned to COG groups with known functions, including 20 gene families involved in ion transport and a series of genes between ISO4G11\_0238 and ISO4G11\_0275 on a single contig, including genes with a toxin and antitoxin annotation.

ISO4-G1 has 526 unique gene families, 261 are predicted to be hypothetical proteins, 82 are poorly characterized, and 196 are assigned to COG groups with known function. These include 34 gene families involved in ion transport, of which 19 gene families were predicted to be involved in iron transport. A non-ribosomal peptide synthase (NRPS) is unique to ISO4-G1 (ISO4G1\_0984), and is predicted to encode a 5,216 aa protein. This NRPS contains two in-frame amber codons, and is predicted to incorporate pyrrolysine. Adjacent to the NRPS gene is an ABC transporter operon (ISO4G1\_0989, ISO4G1\_0990 and ISO4G1\_0991), including an ABC transporter permease (ISO4G1\_0990) that might be involved in export of the NRPS product. There are also three genes encoding a chaperone near the NRPS gene (ISO4G1\_0975, ISO4G1\_0978, ISO4G1\_0979) potentially involved in formation of the NRPS product. Domain analysis of the ISO4-G1 NRPS, shows it contains four modules, each containing an adenylation domain, a condensation domain and ending with a peptidyl carrier protein domain (Figure 4.12). The first module includes an additional adenylation domain and the NRPS ends with an additional condensation domain. The substrate specificity for adenylation domains were predicted from a database of adenylation domains with known substrate, and the predicted

ISO4-G1 NRPS adenylation substrate include ornithine, 2,4-diamino-butyric acid, alanine, tyrosine and pipecolic acid.



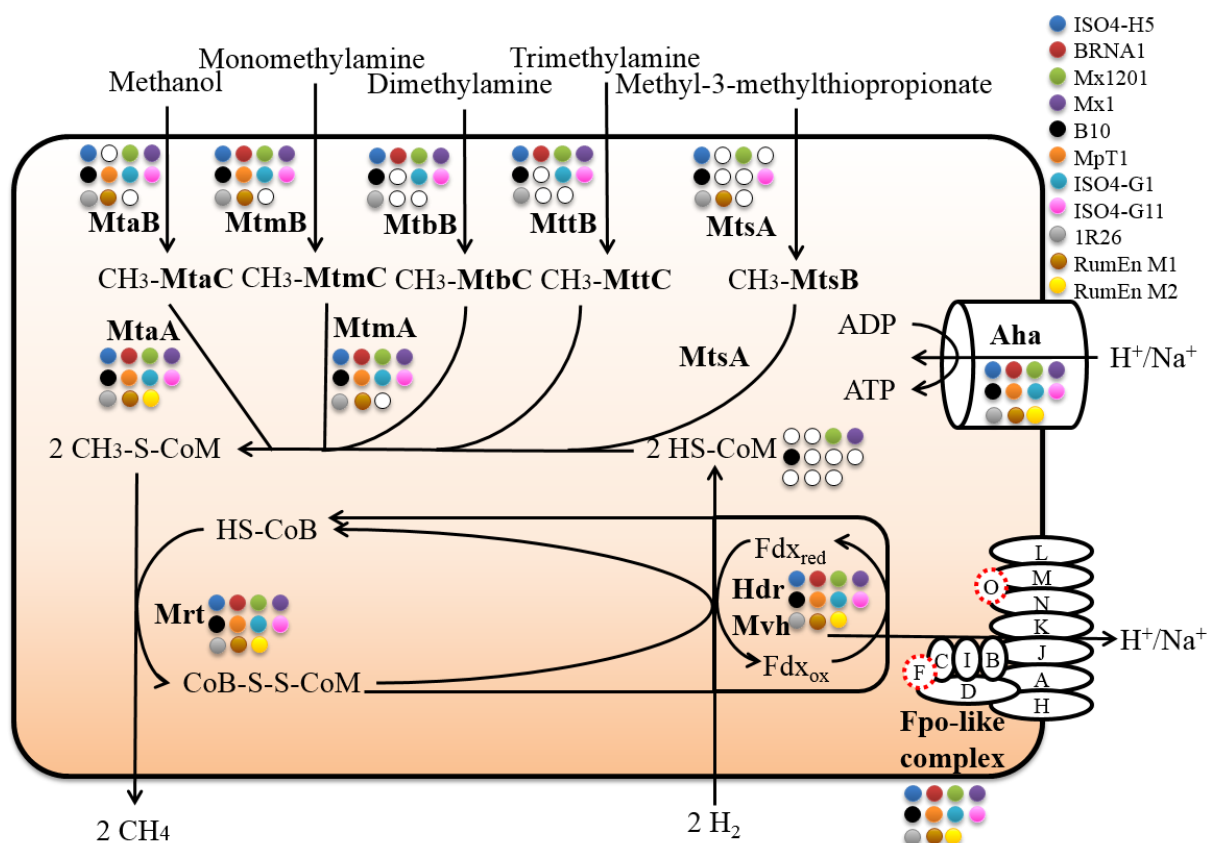
**Figure 4.12 Non-ribosomal peptide synthase of ISO4-G1.** The diagram is not drawn to scale.

#### 4.2.7 Comparative analysis of metabolic profiles predicted from genomic sequences

Comparative analysis of the ORFeomes of the 11 *Methanomassiliicoccales* genomes suggests that the central metabolism and the methanogenesis pathway of these strains are largely similar, with the exception of RumEn M1 genome, which lacks methyl-substrate utilising genes.

*Methanogenesis.* All of the genes predicted to be involved in methanogenesis and energy generation in the genomes analysed are summarised in Table A.4.15. All of the genomes possess a methyl-CoM reductase (*mrt*), a heterodisulfide reductase (*hdr*), and a methyl viologen-dependent hydrogenase (*mvh*) and lack the genes encoding the hydrogenotrophic section of the methanogenesis pathway (Figure 4.13). An Fpo-like complex is predicted to be the key cation transporter linking methanogenesis to the generation of membrane potential, which allows the ATP synthase complex to produce energy driven by a  $\text{Na}^+$  or  $\text{H}^+$  gradient (Table A.4.12).



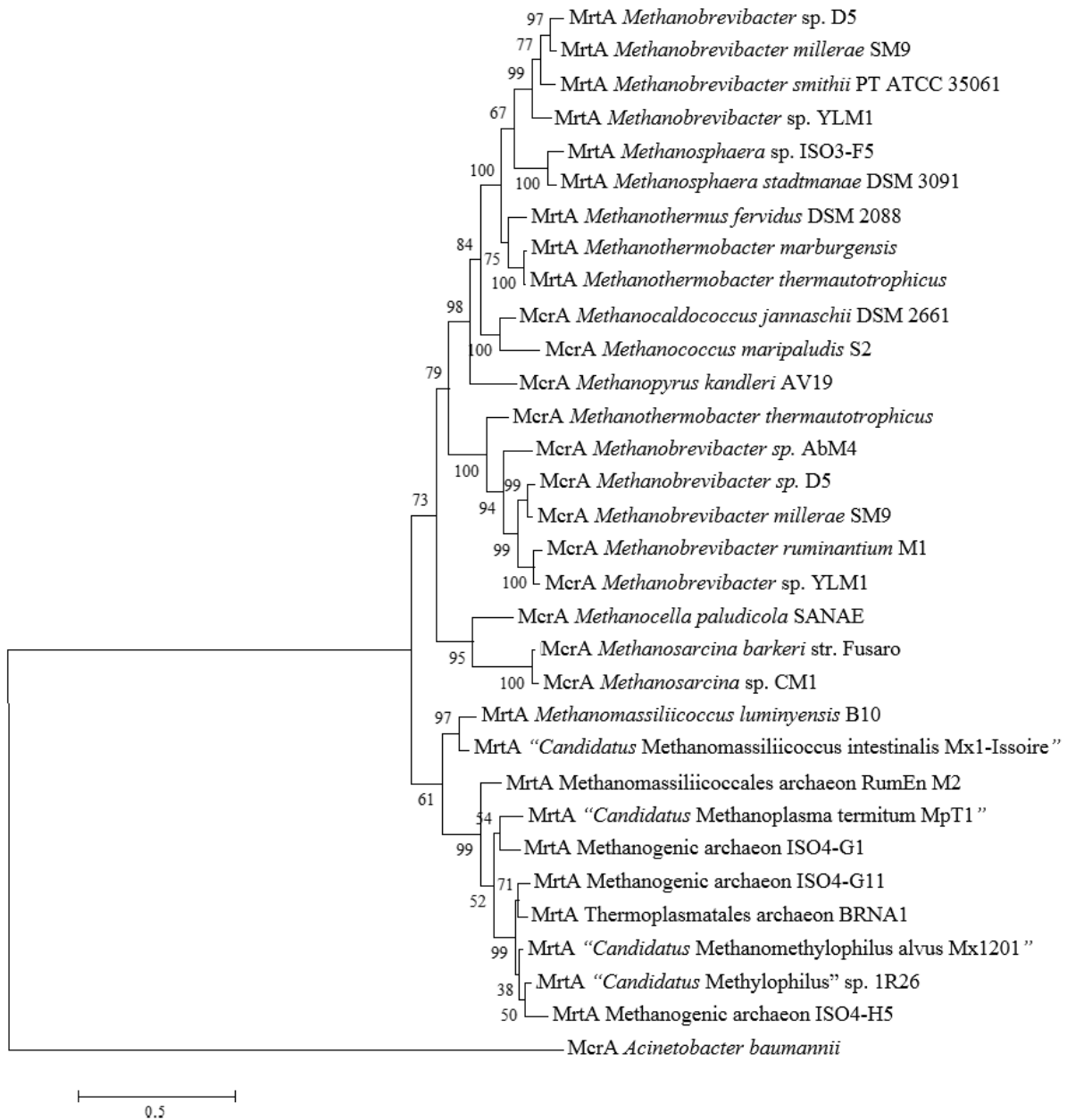


**Figure 4.13 Methanogenesis pathway from methyl compounds by Methanomassiliicoccales.** Key genes are shown in bold. The presence or absence of each gene is highlighted by colored circles. A white circle indicates absence of that gene in the corresponding genome. A red dotted ring indicates the absence of a certain protein subunit within the enzyme complex. Methanol (MeOH), monomethylamine (MMA), dimethylamine (DMA), trimethylamine (TMA), methyl-3-methylthiopropionate (M3MTP), methanol corrinoid protein (MtaC), methyl:coenzyme M methyltransferase (MtaA), methanol:corrinoid methyltransferase (MtaB), trimethylamine:corrinoid methyltransferase (MttB), trimethylamine corrinoid protein (MttC), bifunctional methylthiol:corrinoid methyltransferase (MtsA), methylthiol corrinoid protein (MtsB), monomethylamine:corrinoid methyltransferase (MtmB), monomethylamine corrinoid protein (MtmC), dimethylamine:corrinoid methyltransferase (MtbB), dimethylamine corrinoid protein (MtbC), methyl-coenzyme M reductase (Mrt), methylviologen hydrogenase (Mvh), heterodisulfide reductase (Hdr), ferredoxin (Fdx),  $A_1A_0$  ATP synthase (Aha).

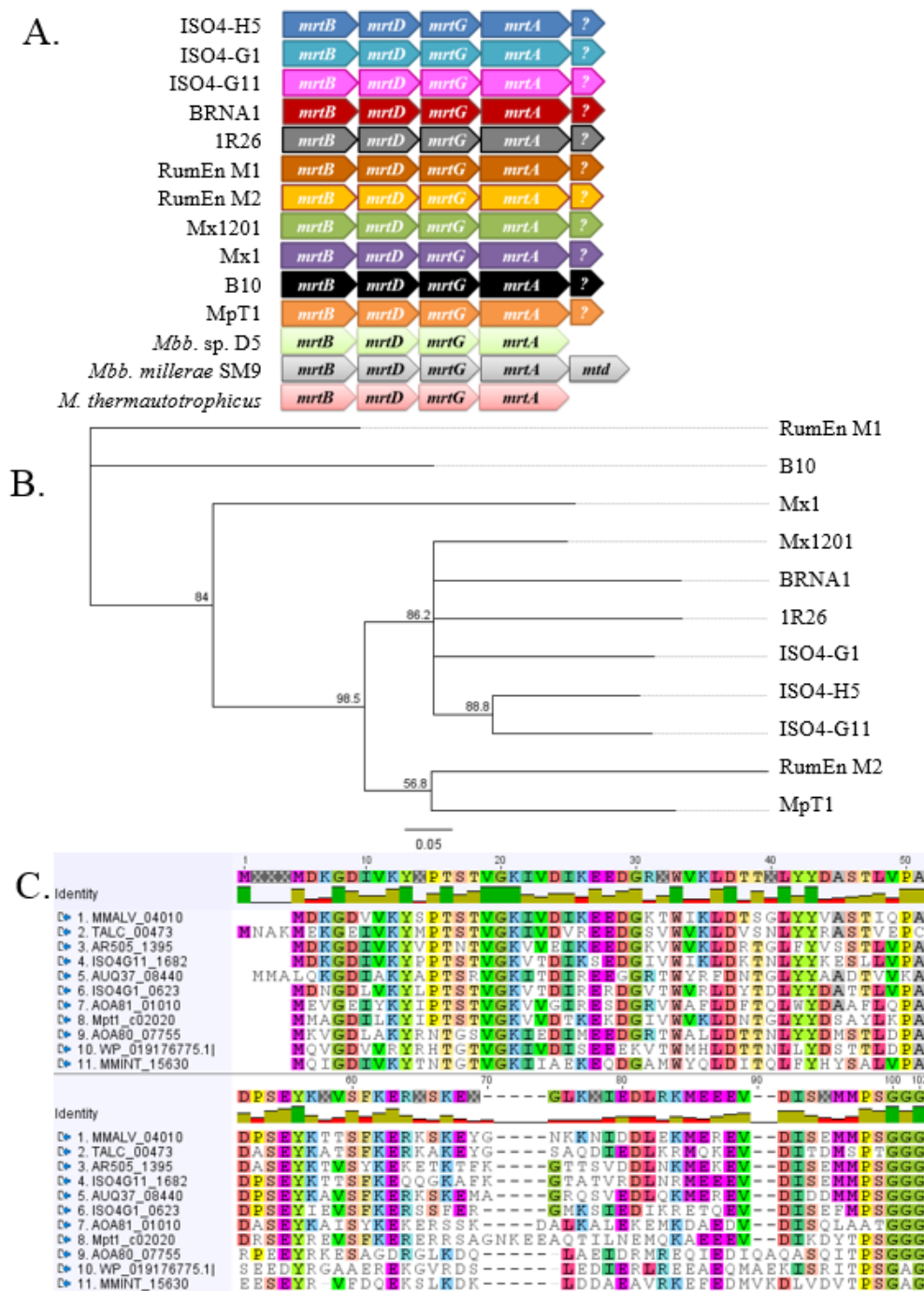
A phylogenetic tree inferred from the amino acid sequence of the McrA/MrtA subunit clearly places members of the order Methanomassiliicoccales in a branch of their own, away from those of Methanosarcinales, Methanobacteriales, and Methanococcales (Figure 4.14). The *mrtG* gene was absent in the 1R26 genome, a frameshift mutation or sequencing error in the stop codon led to its lack of annotation, the *mrtG* gene was identified based on homology and is present between *mrtD* and *mrtA*. The operon organisation of the *mrt* genes of Methanomassiliicoccales has been observed to follow the order of *mrtB*, *mrtD*, *mrtG*, *mrtA* (Figure 4.14A), followed by 3 to 7 other genes, then resumes with *atwA*, *mrtC* downstream on the same strand. Mx1, RumEn M1 and MpT1 are exceptions to this operon organisation; the *atwA* and *mcrC* are located far apart (920 kb apart for Mx1, 14.5 kb apart for RumEn M1 and

427 kb apart for MpT1) from the remainder of the operon (Table A.4.15). The sequence and operon organisation indicate an MCRII/Mrt-type enzyme. In the Methanomassiliicoccales genomes investigated, there is a short conserved hypothetical protein of approximately 92 amino acids, which is located immediately downstream of *mrtA* on the same strand, followed by 2-6 genes which are usually encoded on the opposite strand. This gene is absent from other methanogens that encode *mrt* operons (Figure 4.15A). This hypothetical protein displays a high level of conservation at the aa sequence level ranging from 31.9% aa identity between 1R26 and B10 to 72.8% aa identity between ISO4-H5 and ISO4-G11 (Figure 4.15C). The phylogenetic relationship inferred from the aa sequence of this hypothetical protein is similar to the one that is inferred from the McrA/MrtA sequences (Figure 4.15B), placing RumEn M1, Mx1 and B10 clearly away from other members of the Methanomassiliicoccales order. This hypothetical protein may be functionally associated with the Mrt enzyme complex, but a lack of homology to any known protein precludes assignment of a function.

A number of genes with unknown functions have been identified as genetic markers for methanogens via the TigrFam database, and are known as methanogenesis marker proteins (MMP). Six MMP genes (*mmp3*, *mmp6*, *mmp5*, *mmp15*, *mmp7* and *mmp17*) are located downstream and on the same strand as *mrtC*, whereas *mmp9*, *mmp10*, *mmp12* and *mmp14* are absent amongst the genomes analysed (Table A.4.15).



**Figure 4.14 Phylogenetic tree of methanogens based on McrA/MrtA aa sequences.** The phylogenetic relationships of 26 methanogens were inferred using Maximum Likelihood based on the JTT matrix-based model (Jones *et al.* 1992). Scale bar indicates the number of substitutions per site. The MrtA of RumEn M1 is a pseudogene and is precluded from this analysis. Numbers at the branch point represent bootstrap value from 1000 replicates. *Acinetobacter baumannii* was used as an outgroup.



**Figure 4.15 Conservation of a hypothetical protein associated with the Mrt operon in Methanomassiliococcales.** **A.** *mrt* operon organisation of members of Methanomassiliococcales compared to *mrt* operons of Methanobacteriales; *Methanobrevibacter* sp. D5, *Methanobrevibacter* sp. strain D5; *Methanobrevibacter millerae* SM9 and *Methanobacterium thermautotrophicus*. Methyl-CoM methyltransferase subunit X (*mrtX*, where X = A, B, D or G). Methylene-tetrahydromethanopterin dehydrogenase (*mtd*). Conserved hypothetical protein downstream of *mrtA* gene (?). **B.** Phylogenetic tree inferred from aa sequence of conserved hypothetical protein downstream of *mrtA* gene. Tree generated by Geneious Tree Builder using JTT distance model, built by Neighbour-Joining (Kearse *et al.* 2012). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein 1985) was used to infer the evolutionary history of the taxa analysed. Bar: 0.08 substitutions per amino acid position. **C.** ClustalW alignment of aa sequence of conserved hypothetical protein downstream of *mrtA* gene (Thompson *et al.* 2002), BLOSUM was used as cost matrix 10 gap open cost and 0.1 gap extend cost (Mount 2008).

*Genes encoding methyl compound utilisation.* RumEn M2 genome encodes no genes involved in methyl compound utilisation. Ten members of the Methanomassiliicoccales encode the methylamine:corrinoid methyltransferase MtmB and its corresponding corrinoid protein MtmC. All members of the Methanomassiliicoccales order except BRNA1 and RumEn M2 are predicted to utilise methanol, and all except MpT1 and RumEn M1 to use dimethylamine and trimethylamine (Figure 4.13, Table A.4.15). The copy number of methylamine:corrinoid methyltransferases varies between genomes. B10 carries four copies of *mtmB*, ISO4-G1 carries five copies of trimethylamine:corrinoid methyltransferase (*mttB*). Two *mttB* genes (ISO4G1\_0945 from ISO4-G1 and TALC\_00306 from BRNA1) do not contain an in-frame amber codon.

In addition to methylamine methyltransferases, utilisation of methylamine requires an activation protein for corrinoid proteins and a permease to import the substrates. The methyltransferase corrinoid activation protein (*ramA*) which is required to reactivate cobalt at the active site of MtmC, MtbC and MttC, was found across all members of the Methanomassiliicoccales examined (Table A.4.15), including a RamA homologue in RumEn M2 that shares 43.8% aa identity to experimentally verified RamA of *Methanosarcina barkeri*. There were no monomethylamine permease genes found in any genome, but homologues of dimethylamine permeases were predicted in members of the Methanomassiliicoccales excluding RumEn M1, RumEn M2, MpT1 and B10. Homologues of a trimethylamine permease are similarly distributed in the Methanomassiliicoccales genomes but not in ISO4-G11. No methylamine permease genes could be found at all in the B10 and MpT1 genomes. The bifunctional methylcobalamin:coenzyme M methyltransferase (*mtsA*) and the corresponding corrinoid protein (*mtsB*) required for methylthiol utilisation are only predicted in ISO4-H5, ISO4-G11, RumEn M1, 1R26, Mx1, Mx1201 and B10 (Table A.4.15).

*Heterodisulfide reductase and methyl-viologen hydrogenase.* The cytoplasmic heterodisulfide reductase (*hdrABC*) and methyl-viologen dependent hydrogenase (*mvhAGD*) is highly conserved across all members of the Methanomassiliicoccales order. The *hdrBC* and *mvhAGD* genes are organised as operons, with *mvhD* present in multiple copies in Mx1 and B10 (Table A.4.15). The presence of a lone *hdrD* gene of the periplasmic HdrDE is also conserved across the genomes analysed.

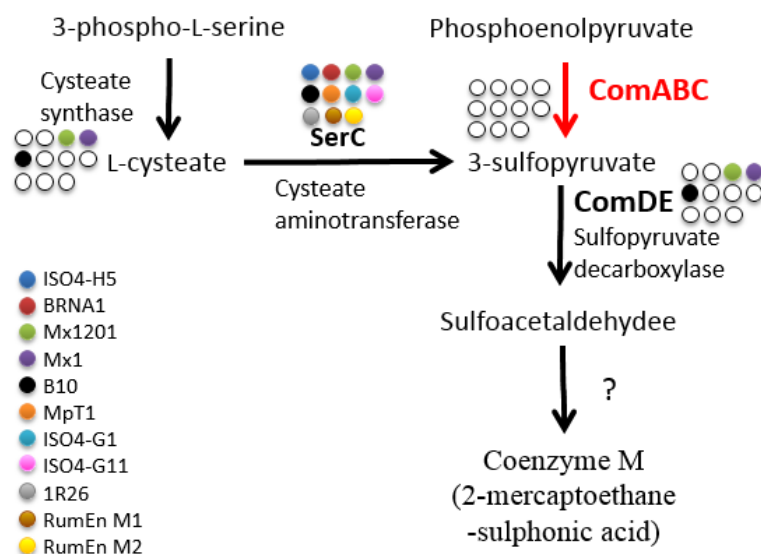
*F<sub>420</sub> methanophenazine oxidoreductase (Fpo)-like complex and hydrogenases.* Members of the Methanomassiliicoccales lack the hydrogenotrophic section of the methanogenesis pathway



Organism	Gene		
RumEn M1	<i>ehaE</i>	VCGTCSNH	DPCICCTDR
B10	<i>ehaE</i>	VCGTCSFHH	DPCICCTDR
Mx1	<i>ehaE</i>	VCGTCSFHH	DPCICCTDR
<i>Methanothermobacter marburgensis</i>	<i>ehaO</i>	VCGTCSYIH	DPCIACAER
<i>Methanothermobacter marburgensis</i>	<i>ehbN</i>	VCGTCSGVH	DPCFTCTDR
<i>Escherichia coli</i>	<i>hycE</i>	VCGTCSFAH	DPCYSCTDR
<i>Escherichia coli</i>	<i>hycG</i>	VCGTCSFAH	DPCYSCTDR
RumEn M2	<i>hypG-like</i>	ISGDNAVAH	GLSYSGNDV
MpT1	<i>hypG-like</i>	ISGDNAIAH	DLSYSGNDL
B10	<i>hypG-like</i>	ISGDTTVAH	NLSYSGNDL

**Figure 4.17** ClustalW alignment of predicted protein amino acid sequence of NiFe binding motif in hydrogenase large subunits. Quality evaluation used BLOSUM as the substitution matrix with 10 gap opening cost and 0.1 gap extension cost (Mount 2008).

*Coenzyme M biosynthesis.* The capability to make CoM varies among the members of the Methanomassiliicoccales. In Mx1201, Mx1 and B10 where it is predicted to be synthesised, it appears not to be made via the typical sulfolactate synthase (*comABC*) pathway, but instead requires cysteate synthase (homologous to threonine synthase), together with a multisubstrate phosphoserine/aspartate aminotransferase (*serC*) followed by a sulfoypyruvate decarboxylase (*comDE*) (Figure 4.18).



**Figure 4.18** Coenzyme M biosynthesis pathway in the Methanomassiliicoccales genomes analysed. Broad substrate specificity phosphoserine transaminase (SerC), (2*R*)-phospho-3-sulfolactate synthase (ComA), 2-phosphosulfolactate phosphohydrolase (ComB), (2*R*)-3-sulfolactate dehydrogenase (ComC), sulfoypyruvate decarboxylase (ComDE), unidentified enzyme (?). Black arrows represent existing pathways. Red arrows represent absent pathways.

A multifunctional phosphoserine/aspartate aminotransferase is found across all members of the Methanomassiliicoccales as it is involved in a wide range of other pathways. However, *comD* and *comE* were only found in Mx1 and B10, and a fused *comDE* was also found in Mx1201. A cysteate synthase homologue has also been found in these three members of the Methanomassiliicoccales. BRNA1 was found to have a phosphonopyruvate decarboxylase



(TALC\_00082) with homology to *comDE* at the 3' end, however, the presence of only a single copy of threonine synthase gene (TALC\_00599) means BRNA1 is unlikely to produce CoM. None of the other members of the Methanomassiliicoccales possess the genes encoding CoM biosynthesis (Table A.4.15).

*Cofactor F<sub>430</sub> biosynthesis.* Cofactor F<sub>430</sub> is used in the active site of the Mcr/Mrt complex but only one gene, *corA* encoding uroporphyrin-III C-methyltransferase is known to be involved in F<sub>430</sub> biosynthesis. All of the Methanomassiliicoccales genomes encode *corA*, except ISO4-G1, ISO4-G11 and RumEn M2 which likely require exogenous sources of F<sub>430</sub> to survive.

*A<sub>1</sub>A<sub>0</sub>-ATP synthase.* Genes encoding nine subunits of A<sub>1</sub>A<sub>0</sub>-ATP synthase are present in all the genomes analysed, and are all positioned close to the putative origins of replication for each organism. The conserved operon structures consist of the genes ordered *ahaHIKECFABD* (Table A.4.15).

*H<sup>+</sup>/Na<sup>+</sup> antiporters.* All the genomes analysed possess H<sup>+</sup>/Na<sup>+</sup> antiporter genes (Table A.4.15).

#### *Central carbon metabolism*

*Gluconeogenesis.* All members of the order Methanomassiliicoccales examined encode the enzymes necessary to carry out gluconeogenesis from pyruvate to glucose-1-phosphate (Figure 4.19, Table A.4.16). However, the enzymes differ and three variants in the gluconeogenesis pathway were found among the genomes analysed. The 2,3-bisphosphoglycerate-independent phosphoglycerate mutase gene (*apgM*) specific to archaea was seen in all genomes while ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, RumEn M1, 1R26 and Mx1201 also possess the bacterial bisphosphoglycerate-dependent *gpmA* variant. B10 was the only genome to possess the phosphofructokinase gene, *pfkB* involved in glycolysis. The MpT1 genome is the only genome lacking the phosphoglucose/phosphomannose isomerase gene, *pgi*.

*Incomplete TCA cycle.* All the genomes are predicted to have an incomplete reductive TCA cycle (Figure 4.19), but the extent of this pathway differs between members of the Methanomassiliicoccales. The ISO4-H5, Mx1, B10, ISO4-G1 and ISO4-G11 possess a citrate synthase gene (*gltA*), which allows the incorporation of acetyl-CoA into the TCA cycle. The B10 genome lacks a recognisable aconitate hydratase gene, *acnA*, but it is the only genome with genes encoding succinate-CoA ligase (*sucCD*) and CoM/CoB dependent thiol:fumarate reductase (*tfrAB*) necessary to interconvert 2-oxoglutarate and fumarate. The ISO4-H5, Mx1201, 1R26 and MpT1 genomes lack genes encoding the 2-oxoglutarate synthase complex



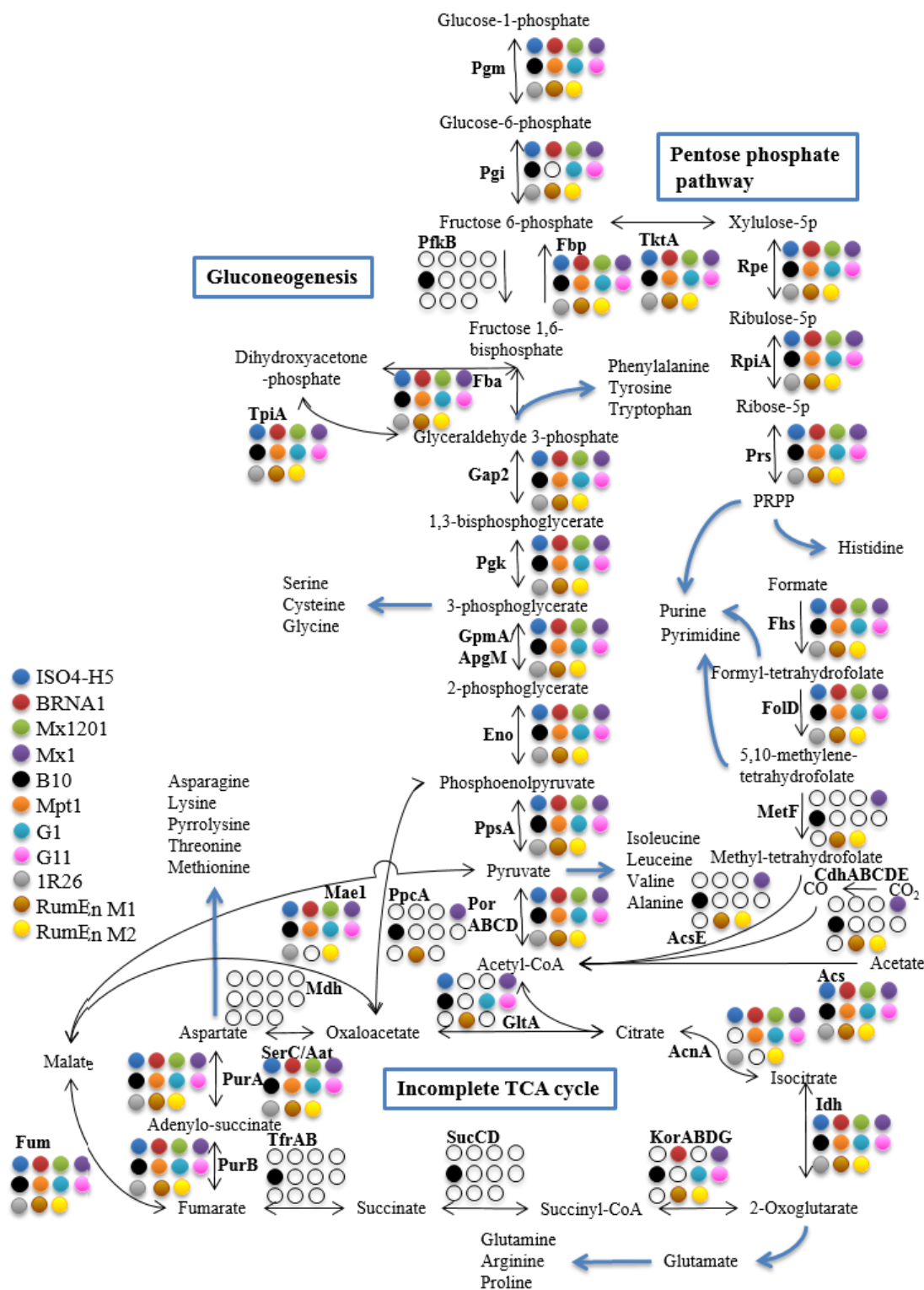
(*korABDG*). A gene for malate dehydrogenase (*mdh*) which interconverts 2-oxaloacetate and malate is absent in all the genomes analysed, but the combined actions of the gene products of adenylosuccinate lyase (*purB*), adenylosuccinate synthetase (*purA*) and the broad substrate specificity phosphoserine aminotransferase (*serC*) can produce oxaloacetate. Oxaloacetate can also be produced by phosphoenolpyruvate carboxylase (*ppcA*), but only the RumEn M1, B10 and Mx1 genomes have the *ppcA* gene.

*Pentose phosphate pathway.* The genes necessary to produce phosphoribosylpyrophosphate (PRPP, used for histidine and purine biosynthesis) from fructose-6-phosphate can be found in all the genomes (Figure 4.19, Table A.4.16), including transketolase (*tktA*), ribulose-phosphate 3-epimerase (*rpe*), ribose-5-phosphate isomerase (*rpiA*) and ribose-phosphate pyrophosphokinase (*prs*).

*Reductive acetyl-CoA pathway.* All members of the order Methanomassiliicoccales analysed are predicted to produce acetyl-CoA from acetate via acetyl-CoA synthetase (*acdA*). While all members of the Methanomassiliicoccales possess a formyltetrahydrofolate synthetase gene (*fhs*) and a bifunctional methenyltetrahydrofolate cyclohydrolase/dehydrogenase (*fold*) to produce 5,10 methylene-tetrahydrofolate (5,10-mTHF, an important cofactor found with many enzymes) from formate, only Mx1 and B10 encode an incomplete carbon monoxide dehydrogenase complex (*cdhABCDE*) granting the capability of *de novo* biosynthesis of acetyl-CoA from formate and CO<sub>2</sub> (Figure 4.19, Table A.4.16).

*Alternative carbon sources.* Members of the Methanomassiliicoccales may utilise alcohol as an alternative carbon source. A type IV iron containing alcohol dehydrogenase (*adh*) is predicted in ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, 1R26 and Mx1201 (Table A.4.16) which is homologous to a multifunctional NADP-dependent alcohol dehydrogenase (NC\_001988.2) in *Clostridium*, capable of producing acetyl-CoA from ethanol. A homologue to a cytoplasmic D-lactate dehydrogenase (*ldhA*) is also predicted in BRNA1, RumEn M1, RumEn M2, 1R26 and Mx1201, which may convert lactate to pyruvate.

*Carbon fixation.* Homologues of ribulose 1,5-bisphosphate carboxylase (*rbcL*) and ribulose 1,5-bisphosphate isomerase (*e2b2*) have been identified among the core genome of Methanomassiliicoccales. In all eight genomes analysed, *rbcL* and *e2b2* form a conserved operon with AMP phosphorylase (*deoA*).



**Figure 4.19 Central carbon metabolism pathways predicted for Methanomassiliicoccales.** Key genes are shown in bold. Black arrows indicate pathways, blue arrows represents pathways leading out of central carbon metabolism. The presence or absence of each gene in a particular genome is highlighted by colored circles: a white circle indicates absence from a genome, and a coloured circle indicates the presence in the corresponding genome. Formate-tetrahydrofolate ligase (Fhs), NADP-dependent methylene tetrahydrofolate dehydrogenase (FoID), 5,10-methylenetetrahydrofolate reductase (MetF), carbon-monoxide dehydrogenase complex (CdhABCDE), methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase (AcsE), acetyl CoA synthetase (Acs), phosphoglucomutase (Pgm), phosphoglucose isomerase (Pgi), fructose-1,6-bisphosphatase (Fbp), phosphofructokinase (PfkB), fructose-bisphosphate aldolase (Fba), glyceraldehyde-3-phosphate dehydrogenase (Gap2), phosphoglycerate kinase (Pkg), phosphoglycerate mutase (ApgM), phosphopyruvate



All members of order Methanomassiliicoccales analysed encodes genes required to produce alanine, asparagine, aspartate, cysteine, glutamine, histidine, lysine, proline, pyrrolysine and threonine (Table A.4.17, Table 4.10).

**Table 4.10 Amino acid biosynthesis summary**

Amino acids	ISO4-H5	ISO4-G1	ISO4-G11	BRNA 1	RumEn M1	RumEn M2	1R26	Mx120 1	Mx1	B10	MpT1
Alanine	+	+	+	+	+	+	+	+	+	+	+
Arginine	+	+	+	+	+	+	+	+	+	-	+
Asparagine	+	+	+	+	-	+	+	+	+	+	+
Aspartate	+	+	+	+	+	+	+	+	+	+	+
Cysteine	+	+	+	+	+	+	+	+	+	+	+
Glutamine	+	+	+	+	+	+	+	+	+	+	+
Glutamate	+	+	+	+	+	-	+	+	-	+	-
Glycine	-	+	+	+	+	+	-	+	+	+	+
Histidine	+	+	+	+	+	+	+	+	+	+	+
Isoleucine	+	-	-	-	-	+	+	+	+	+	+
Leucine	+	-	-	-	-	+	+	+	+	+	+
Lysine	+	+	+	+	+	+	+	+	+	+	+
Methionine	-	-	-	-	-	-	-	+	-	-	-
Phenylalanine	+	+	+	+	+	+	+	-	-	+	+
Proline	+	+	+	+	+	+	+	+	+	+	+
Pyrrolysine	+	+	+	+	+	+	-	+	+	+	+
Serine	+	+	+	+	+	+	+	+	+	-	+
Threonine	+	+	+	+	-	+	+	+	+	+	+
Tryptophan	+	-	+	+	-	-	+	+	-	+	+
Tyrosine	+	+	+	+	+	+	+	+	+	-	+
Valine	+	-	-	-	-	+	+	+	+	+	+

+: This genome is predicted to be able to produce this amino acid. -: This genome is predicted to be unable to produce this amino acid.

Several genomes possess incomplete pathways for amino acid biosynthesis. Mx1 and MpT1 lack a glutamate dehydrogenase gene (*gdhA*) required to produce glutamate. B10 lacks the acetylglutamate kinase gene (*argB*) for arginine biosynthesis, *serAB* genes required for serine biosynthesis, shikimate dehydrogenase (*aroE*) required to produce chorismate, and prephenate dehydrogenase (*tyrA*) required to produce tyrosine. ISO4-H5 lacks the serine hydroxymethyltransferase (*glyA*) required to convert serine to glycine. Mx1201 and Mx1 lack the broad substrate aromatic aminotransferase (*aro8*) involved in interconversion between aromatic amino acids. The isopropylmalate/citramalate isomerase *leuCD* genes involved in branched chain amino acid biosynthesis is predicted to be absent in the ISO4-G1, ISO4-G11 and BRNA1 genomes. Furthermore, ISO4-G1 lacks the acetolactate synthase (*ilvBN*) and dihydroxy-acid dehydratase (*ilvD*) gene. The genes required for tryptophan biosynthesis are entirely missing in the ISO4-G1 genome, while Mx1 encodes only tryptophan synthase beta chain (*trpB*) and anthranilate synthase (*trpD*) genes (Table A.4.17).

Only the Mx1201 genome encodes a complete methionine biosynthesis pathway (Table A.4.17). While the *metA* (homoserine *O*-succinyltransferase) and *oah* (*O*-acetyl-L-homoserine sulfhydrylase) genes were identified in members of the Methanomassiliicoccales analysed except RumEn M1, RumEn M2 and B10, the other genes encoding the pathway are present

sporadically in the genomes. All the genes necessary to carry out the S-adenosyl-L-methionine (SAM) cycle, apart from methionine synthase itself, have been identified in the ISO4-H5 and Mx1 genomes. All the genomes analysed possess the *speA* gene involved in agmatine biosynthesis and lack the *speE* gene involved in spermidine biosynthesis.

*Nucleotide biosynthesis.* All the genes required for purine and pyrimidine biosynthesis have been identified across members of the Methanomassiliicoccales analysed (Table A.4.18; Figure 3.12 also shows a schematic for the pathway of nucleotide biosynthesis). The only exception is the *ndrJ* gene which is absent from the RumEn M1, RumEn M2 and B10 genomes.

*Cell cycle.* The genes necessary for cellular replication are summarised in Table A.4.19. The B10 draft genome lacks genes encoding a replication factor C large subunit, a DNA primase large subunit and a DNA gyrase subunit A (*gyrA*) as well as a cell division GTPase. All members of the Methanomassiliicoccales examined have two copies of the *orc1/cdc6* origin of replication genes.

*Cofactor biosynthesis.* Tetrapyrrole is the backbone of cobalamin and vitamin cofactors, including F<sub>430</sub>. Members of the Methanomassiliicoccales produce tetrapyrrole by using the glutamyl-tRNA via an uroporphyrinogen-III intermediate, but are incapable of producing cytochromes. The entire set of genes involved in tetrapyrrole biosynthesis is absent in the ISO4-G1 genome, and the ISO4-G11 genome only has the porphobilinogen deaminase (*hemC*) gene, RumEn M2 genome only has the siroheme synthase (*cysG*) gene (Table A.4.20), therefore ISO4-G1, ISO4-G11 and RumEn M2 are predicted not to produce cofactor F<sub>430</sub>. Several genes required for adenylosylcobalamin biosynthesis are predicted to be absent in RumEn M1 and RumEn M2 genomes (Table A.4.20), therefore RumEn M1 and RumEn M2 are predicted not to carry out *de novo* biosynthesis of adenylocobalamin.

All the Methanomassiliicoccales genomes analysed are predicted to possess the genes required for riboflavin biosynthesis, as well as subsequent FAD and FMN biosynthesis (Table 4.19). Two genes are considered to be variants of the bacterial gene rather than the archaeal gene, the bifunctional diaminohydroxy phosphoribosylaminopyrimidine deaminase /5-amino-6-(5-phosphoribosylamino) uracil reductase (*ribD*) and the bacterial ATP-driven riboflavin kinase (*ribC*). The archaeal *ribC* gene product is CTP-dependent.

The genes required for *de novo* biosynthesis of NAD<sup>+</sup> and NADP<sup>+</sup> were only found in the B10 genome. The other genomes lack the L-aspartate oxidase gene (*nadB*) which encodes the first

step of NAD<sup>+</sup> biosynthesis. All the genomes analysed possess the genes required for salvaging of NAD and NADP (Table A.4.20).

*Secretion.* All the Methanomassiliicoccales genomes analysed are predicted to possess genes required for protein secretion (Table A.4.21). The secretion is signal recognition particle dependent, the signal recognition particle is recognized by signal Signal recognition receptor FtsY and cleaved by signal sequence peptidase, then exported via the Sec61 translocon.

### 4.3 Discussion

Eleven genomes from members of the order Methanomassiliicoccales sourced from four different gut environments (human, termite, bovine rumen, and ovine rumen) have been compared in this study, to identify the extent of gene conservation and to gain insight into the metabolism and physiology of this unusual group of methanogens. The genomes analysed all contain circular chromosomes with no extrachromosomal elements. All genomes are complete with the exception of B10, RumEn M1 (96.6% completion), RumEn M2 (94.1% completion) (Sollinger *et al.* 2016), 1R26 (92.98% complete) (Noel *et al.* 2016) and G11 which are currently in draft form. Amongst the sequenced members of the Methanomassiliicoccales, the genome size ranges between 1.46 Mbp (BRNA1) and 2.62 Mbp (B10), and the % G+C range is between 41.3% (Mx1) and 62.1% (RumEn M1). A closer examination of those strains isolated from the rumen reveals a genome size of between 1.46 Mbp (BRNA1) and 1.94 Mbp (ISO4-H5), and a % G+C range of between 54.0% (ISO4-H5) and 58.3% (BRNA1). Members of the Methanomassiliicoccales order were originally associated with the Thermoplasmatales order (Tajima *et al.* 2001), whose genomes are known to lack an operon organisation of their rRNAs (Tu and Zillig 1982). Separation of the 16S, 23S and 5S rRNA genes has been reported in genomes of the Methanomassiliicoccales isolated from human and termite gut environments (Borrel *et al.* 2014; Lang *et al.* 2015). Analysis of the rRNA gene structure in the genomes isolated from the rumen environment confirms this separation of the rRNA genes, and appears to be a genomic trait shared by all members of the Methanomassiliicoccales.

The phylogenetic relationship of members of the Methanomassiliicoccales can be inferred from the 16S rRNA gene sequence (Seedorf *et al.* 2014). The relationship inferred from the 16S rRNA gene suggests that the organisms B10 and Mx1, both isolated from human faeces, and RumEn M1 isolated from bovine rumen are phylogenetically distant to the other sequenced members of the Methanomassiliicoccales. These three organisms group more closely with the environmental isolates of Methanomassiliicoccales (Figure 4.1B), as reported previously (Sollinger *et al.* 2016). A FGD analysis reveals a similar profile (Figure 4.1A) whereby the RumEn M1, B10 and Mx1 genomes group together in a gastrointestinal-tract cluster (GIT2) away from the ISO4-H5, BRNA1, 1R26, RumEn M2, ISO4-G11, Mx1201 and ISO4-G1 genomes in GIT1. The conservation in genome structure and gene order between different genomes is known as synteny (Passarge *et al.* 1999). When two genomes have a high degree of synteny, a linear genome alignment can be observed across the entire length of the genome,

and when an X-alignment is observed, as reported for *Mycobacterium leprae* and *Mycobacterium tuberculosis*, this indicates that matching sequences occur at the same distance from the origin but not necessarily on the same side of the origin (Eisen *et al.* 2000). This is often seen in synteny studies of moderately diverged genomes (Eisen *et al.* 2000). Using the ISO4-H5 genome as a reference, it was shown that a high degree of synteny is shared with the BRNA1 and Mx1201 genomes. A X-alignment was observed in these analysis and the identified syntenic breakpoints included a TPR repeat-containing protein (AR505\_0536) and a region downstream of a an ATP-dependent DNA helicase (AR505\_1369) and upstream of a ribosomal protein S27e (AR505\_1370). This large-scale inversion encompasses 955 kb of the genome.

The codon usage of the genomes was compared and found to be similar between all members of the Methanomassiliicoccales, with one exception being the Mx1 genome. Mx1 was found to have a low G+C% (41.3%) content, and this was reflected in the codon usage of the amino acids, phenylalanine, asparagine, lysine, glutamate, tyrosine, alanine and cysteine, where Mx1 preferentially uses codons with less G or C content (Table A.4.1). Despite this observed difference in codon usage, all of the Methanomassiliicoccales genomes maintain a conserved amino acid usage pattern. An analysis of the genome nucleotide similarity using the ISO4-H5 genome as a reference, highlighted the presence of a 25,429 bp region with 99.96% sequence identity and a 22,621 bp region with 99.26% sequence identity to the draft genome sequence of ISO4-G11. Eight adhesin-like protein genes, six ABC transporter-related genes, an ATP-dependent DNA helicase gene and 15 hypothetical protein genes are predicted within these two regions. This high level of sequence identity was not observed between other Methanomassiliicoccales genomes analysed. The ISO4-G11 genome was assembled from an enrichment culture and is currently in draft form, as such, the presense of such regions will need to be investigated further in the ISO4-G11 assembly.

All of the genomes analysed are predicted to encode a full set of tRNAs corresponding to 21 amino acids, except the incomplete RumEn M2 that lacks tRNA<sup>Pro</sup>. The tRNAs which contain introns differ between genomes. Intron-containing tRNAs are found in all three kingdoms of life (Yoshihisa 2014). The RumEn M1, RumEn M2, Mx1 and B10 genomes are predicted to possess a high number of intron containing tRNAs (eight, nine, seven and eight respectively), with conserved intron-containing tRNAs (tRNA<sup>Arg</sup>, tRNA<sup>Asn</sup>, tRNA<sup>Cys</sup> and tRNA<sup>Ser</sup>), while the other strains possess a low number of intron containing tRNAs (three or four) with conserved intron containing tRNA<sup>Met</sup>. All strains contain a conserved set of two intron-containing tRNAs



(tRNA<sup>Trp</sup> and tRNA<sup>Tyr</sup>). While Mx1201 is closely related to the genomes of ovine rumen origin based on 16S rRNA gene inferred phylogeny, the MpT1 genome is more distantly related. The identical intron-containing tRNAs between MpT1 and Mx1201 suggests that both phylogeny and environment could contribute to the distribution of intron-containing tRNAs. RumEn M2 is clustered in GIT1 based on 16S rRNA gene inferred phylogeny, the distribution of intron-containing tRNAs is highly similar to RumEn M1, B10 and Mx1 of GIT2. This suggests RumEn M2 may belong to a currently under-represented clade of Methanomassiliicoccales and differ to other members of GIT1.

All except one genome of the Methanomassiliicoccales examined are predicted to encode and utilise the 22<sup>nd</sup> aa, pyrrolysine. The incomplete RumEn M2 genome is predicted to lack the pyrrolysine biosynthesis genes and all methyltransferases that incorporates pyrrolysine. The biosynthetic machinery of pyrrolysine is organized as a highly conserved operon (Gaston *et al.* 2011; Gaston *et al.* 2011). This operon structure is conserved among the genomes analysed, along with the surrounding methylamine utilisation genes (Figure 4.4). This co-localisation is important because pyrrolysine is thought to be required for the catalytic activity of MtmB, MtbB and MttB (Krzycki 2004). The genes containing in-frame amber codon(s) have been reported for the genomes of B10, Mx1 and Mx1201 (Borrel *et al.* 2014). In this study, the in-frame amber-containing genes of the 11 genomes were examined and then classified into three categories: Class 1 genes which have amber codon read-through and subsequent incorporation of pyrrolysine; Class 2 genes that utilised the amber codon as a stop codon; Class 3 genes with uncertain amber codon usage due to lack of homologous genes. From this investigation it is clear there is a distinct difference in the predicted pyrrolysine incorporation amongst the genomes analysed (Figure 4.5). The genomes of RumEn M1, RumEn M2, B10 and Mx1 encode a high number of genes (13%, 8.2%, 11.3% and 5.3% respectively) with an in-frame amber codon (Table 4.4) and a small number of Class 1 genes predicted to incorporate pyrrolysine. This is in comparison to the other genomes who have a lower number of genes (1.8% to 4.1%) predicted to have an in-frame amber codon (Table 4.4), but a higher number of Class 1 genes predicted to incorporate pyrrolysine (Figure 4.5). Based on this analysis, two separate genomic profiles of pyrrolysine incorporation may exist within the gut environment and correspond to the two GIT clusters identified based on FGD analysis, with the exception of RumEn M2. The RumEn M2 genome contained no pyrrolysine biosynthesis genes, however, its genome is only 94.2% completion, it is likely the pyrrolysine biosynthesis genes are absent from the currently sequenced genome, but may be present once the genome is completed and

circularised. This assumption is supported by two factors in addition to the predicted usage of amber codon. i) The absence of any methyl-substrate utilising methyltransferase suggests a vital part of the genome is missing, the current genome indicates it is incapable of carrying out all known methanogenesis pathways. The methylamine methyltransferase genes are typically closely located with the pyrrolysine biosynthesis genes, therefore both the methylamine utilising genes and the pyrrolysine biosynthesis genes may present itself once the genome is completed. ii) A RamA homologue was predicted in the RumEn M2 genome, which is required to activate the corronid protein utilised by the methylamine methyltransferases, therefore this study analysed RumEn M2 genome in assumption that the full genome would contain a pyrrolysine biosynthesis operon.

For the Class 2 genes predicted to use the amber codon as a stop codon (10, 6, 5, 6, 37, 26 and 13 genes for ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, 1R26, Mx1201 and MpT1 respectively), the distance between the amber codon and the next opal/ochre stop codon is usually short, suggesting the additional 3' amino acids are unlikely to affect the function of the gene product (Table 3.6, Table A.4.2, Table A.4.03, Table A.4.4, Table A.4.5, Table A.4.8). These results indicate that even if every predicted gene with an in-frame amber codon incorporates pyrrolysine due to a lack of regulation, the genomes of ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, 1R26 and Mx1201 are unlikely to suffer any adverse consequences, as the Class 2 genes in these genomes are predicted to be non-essential. For example, in the ISO4-H5 genome, nine IS elements and one adenylate kinase gene are predicted to utilise the amber codon as a stop codon. The absence of any pyrrolysine regulation in the ISO4-H5 organism would allow production of an extended protein past the amber codon, which may be misfolded and dysfunctional. The production of dysfunctional IS elements and an adenylate kinase may cost the ISO4-H5 organism some energy, but a second copy of an adenylate kinase gene in the ISO4-H5 genome is likely to keep the metabolism of ISO4-H5 functionally intact. For organisms such as B10 and Mx1, it has been proposed that they are able to distinguish the genes that require pyrrolysine to function, and as such, selectively incorporate the pyrrolysine, whereas all other genes with an in-frame amber codon are regulated to use the amber codon as a stop codon (Borrel *et al.* 2014). This analysis suggests that some members of the Methanomassiliicoccales may employ a mechanism of regulation of amber-codon suppression, whereas others do not, but rely on constitutive expression of the pyrrolysine biosynthesis operon.

This difference in regulation is likely related to the phylogenetic divergence between members of the Methanomassiliicoccales as inferred from FGD analysis with RumEn M1, B10 and Mx1 clustered separately from the remaining genomes (Figure 4.1A). The availability of methylamine substrates may also be linked to this difference in regulation. For example, the Mx1201 isolate is proposed to have evolved in an environment where methylamines are not limiting, and alongside methylamine producing microorganisms (Smith and Macfarlane 1996; Borrel *et al.* 2014). Colonic anaerobes of the genera *Clostridium*, *Bifidobacterium* and *Bacteriodes* have been shown to utilise choline and glycine betaine, producing methylamines and propylamines in the presence of fermentable carbohydrates (Allison and Macfarlane 1989; Smith and Macfarlane 1996). The high availability of methylamines may stimulate the constitutive expression of the pyrrolysine operon as Mx1201 evolved, and led to the low usage of the amber codon as a stop codon (Borrel *et al.* 2014). Similarly, methylamines are readily available in the rumen, as choline from plant sap (Maizel *et al.* 1956), cell membrane (Neill *et al.*, 1978) and betaine produced from chloroplasts (Rhodes and Hanson 1993) and are metabolised in the rumen (Neill *et al.* 1978; Mitchell *et al.* 1979). The availability of methylamines may have contributed to the routine usage of pyrrolysine in rumen members of the Methanomassiliicoccales. The Mx1201 genome is known to contain essential genes requiring amber codon read-through to fully translate proteins (Borrel *et al.* 2014), which indicated that the constitutive expression of the pyrrolysine operon and pyrrolysine is essential for growth in Mx1201 (Borrel *et al.* 2014). Similar genes have been found in ISO4-H5, ISO4-G1, ISO4-G11 and BRNA1, such as a bifunctional phosphoglucose/phosphomannose isomerase (AR505\_0560), aconitate hydratase (ISO4G1\_0572), serine hydroxymethyltransferase (ISO4G11\_0237) and carbamoylphosphate synthase (TALC\_00554). All of these genes require pyrrolysine incorporation, which suggests the pyrrolysine operon may be constitutively expressed and essential for growth in these members of Methanomassiliicoccales.

Although the MpT1 and 1R26 genomes display a genomic profile of pyrrolysine incorporation akin to what is observed for ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, Mx1201, it does not possess genes encoding essential functions, which require amber codon read-through to be fully translated. The MpT1 and 1R26 genes with an in-frame amber codon are either hypothetical proteins, mobile elements or genes with a very short distance between the amber codon and the next stop codon, which suggests that MpT1 and 1R26 may not suffer adverse consequences without pyrrolysine usage regulation. The pyrrolysine usage profile of the MpT1 and 1R26 genome places themselves between the group that regulates pyrrolysine usage, (RumEn M1,

B10, Mx1) and the group that minimizes in-frame amber codons (Mx1201, BRNA1, ISO4-H5, ISO4-G1 and ISO4-G11), which correlates with the phylogenetic tree inferred from 16S rRNA genes (Figure 4.1, Figure 4.5).

A variety of pyrrolysine regulation mechanisms have been observed previously, including substrate regulated pyrrolysine operon expression in *Acetohalobium arabaticum* and *Methanosarcina acetivorans* (Mahapatra *et al.* 2006; Prat *et al.* 2012), and usage of different release factors in *M. barkeri* and *M. acetivorans* (Alkalaeva *et al.* 2009). The presence of a pyrrolysine insertion sequence within the *mtmB* gene has also been shown to regulate pyrrolysine usage in *M. barkeri* at the transcription level (Longstaff *et al.* 2007). *Acetohalobium arabaticum* has been observed to regulate pyrrolysine incorporation by presence of a substrate (Prat *et al.* 2012). It is unclear at this time whether regulation methods such as these operate within members of the Methanomassiliicoccales analysed.

Extracellular and surface-associated proteins are likely to play an important role in many essential interactions and adaptations of members of the Methanomassiliicoccales to their environment. In the Methanomassiliicoccales genomes analysed, an average of 4.3% of their ORFeomes are devoted to the process of protein export. ISO4-H5 is predicted to have the highest proportion of exported gene products, 5.6% of its ORFeome, while RumEn M2 is predicted to have the lowest proportion of exported gene products, 2.5% of its ORFeome. Proteins predicted to contain a signal peptide without a transmembrane helice or lipobox are thought to be secreted into the environment. The number of proteins predicted to be secreted into the environment range between 4 and 25 per genome for the genomes analysed. Extracellular proteins with domain repeats are often involved in binding to the cell (Cabanes *et al.* 2002), whereas those that contain a transglutaminase domain may be able to mediate adhesion to a binding partner (Ng *et al.* 2016). All genomes examined contain genes with predicted repeat domains (17, 6, 19, 9, 4, 4, 10, 8, 4, 27 and 10 genes for ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, RumEn M1, RumEn M2, 1R26, MpT1, Mx1201, Mx1 and B10 respectively), while genes containing a transglutaminase-like domain have been predicted in eight of the genomes, ISO4-G1, ISO4-G11, BRNA1, RumEn M1, 1R26, MpT1, Mx1201 and MpT1. These genes suggest that members of the Methanomassiliicoccales are capable of binding to other microorganisms or their hosts, similar to how *Mbb. ruminantium* M1<sup>T</sup> binds to protozoa (Ng *et al.* 2016). Amongst the secretome, there are two transporters with Pfam domains conserved but not limited to six rumen members of the Methanomassiliicoccales. These domains are the PF00924 mechanosensitive ion channel and PF01032 FecCD transport

family. The mechanosensitive ion channels are pressure sensitive ion channels involved in osmoregulation according to environmental conditions (Kloda and Martinac 2001), which may be an important adaptation to the rumen environment. Such transporters may help members of the Methanomassiliicoccales, who lack a cell wall structure, to cope with the osmotic pressure of the outside environment. The FecCD transport family is involved in iron transport (Staudenmaier *et al.* 1989). Methanogens depend on electron transport chain to conserve energy from methanogenesis, Fe-S centres are present in Hdr, ferredoxins involved in methanogenesis and other enzymes of various metabolic functions, therefore an efficient iron uptake system may be beneficial to the members of the order Methanomassiliicoccales. Iron has been found to be acting as electron donor for methanogenesis in *Methanococcus maripaludis* KA1 (Uchiyama *et al.* 2010), however *M. maripaludis* KA1 utilises possess the hydrogenotrophic methanogenesis pathway that is absent in members of Methanomassiliicoccales, and therefore it is unlikely for Methanomassiliicoccales to utilise iron as an electron donor. The attempt to verify secretome predictions by a proteomic approach in three *Sulfolobus* spp. have revealed 64 proteins amongst the supernatant and cell surface subproteome, with cell surface proteins dominating the secreted protein, furthermore, the majority of secreted proteins originated from the cell surface (Ellen *et al.* 2010). Which suggests genuine protein secretion is a rare event in archaea, and experimental validation is required to assess predicted secreted proteins and their biological role.

Rumen methanogens are known to utilise hydrogenotrophic, acetoclastic and methylotrophic methanogenesis pathways to produce energy (Thauer 1998). The methylotrophic methanogens within the rumen environment include *Methanosarcina* spp., *Methanosphaera* spp. and members of the Methanomassiliicoccales. The *Methanosarcina* spp. are versatile methanogens containing cytochromes (Kühn *et al.* 1983) and methanophenazine (Abken *et al.* 1998), and rely on both a membrane bound and cytoplasmic heterodisulfide reductase (Buan and Metcalf 2010) in addition to a proton pump for energy conservation (Baumer *et al.* 2000). They are capable of carrying out hydrogenotrophic methanogenesis from H<sub>2</sub> and CO<sub>2</sub>, acetoclastic methanogenesis by disproportionating acetate (Fournier and Gogarten 2008), and methylotrophic methanogenesis by disproportionating methanol (Muller *et al.* 1986), where the reducing potential is derived from oxidising one mole of methanol to reduce three moles of methanol to methane (Thauer *et al.* 2008). *Methanosphaera* spp. are methylotrophic methanogens that possess the pathway of hydrogenotrophic methanogenesis, yet rely on methanol and H<sub>2</sub> for methane production (Fricke *et al.* 2006). *Methanosphaera stadtmanae* has

been shown to be incapable of molybdopterin biosynthesis, which is required for an active formylmethanofuran dehydrogenase in hydrogenotrophic methanogenesis, making it unable to carry out hydrogenotrophic methanogenesis or to carry out disproportionation of methanol (Fricke *et al.* 2006). Members of the Methanomassiliicoccales are specialised in direct reduction of methylated compounds for methylotrophic methanogenesis, and are unable to disproportionate the methylated substrates like members of the *Methanosarcina* (Welanders and Metcalf 2005) (Figure 4.13). Members of the Methanomassiliicoccales distinguish themselves from *Methanosphaera* spp. by the absence of the cofactor F<sub>420</sub> (E<sub>0</sub>' = -360 mV), and the use of ferredoxin (-500 mV) as an electron carrier, which is more exergonic (Welte and Deppenmeier 2014; Lang *et al.* 2015), furthermore, the Hdr/Mvh are not predicted to be coupled to an energy converting-hydrogenase complex (Thauer *et al.* 2008), but rather are coupled to a F<sub>420</sub>-dehydrogenase Fpo-like complex. As such, these variations in the methylotrophic methanogenesis pathway has allowed *Methanosarcina* spp. (0.02% abundance in the rumen environment), members of the Methanomassiliicoccales (10.4%), and *Methanosphaera* spp. (13.8%) (Seedorf *et al.* 2015) to co-exist and to find their particular niche within the rumen environment. Interestingly, the Tibetan yak was found to have 80.9% of its rumen methanogen population as members of the Methanomassiliicoccales (Huang *et al.* 2012), which could possibly be contributed by diet. This particular study sampled yaks that were fed on *Kobresia* pasture (Huang *et al.* 2012), a dominant genera of plants in the alpine meadow. *Kobresia tibetica* Maximowicz was found to contain high level of condensed tannin (4.79% dry matter) (Niu 2014). Tannin has been proposed to be inhibitory to methanogenesis (Goel and Makkar 2012), potentially through inhibition of hydrogen-producing protozoa (Makkar *et al.* 1995). In *Bos indicus*, treatment with condensed tannin increases the Methanomassiliicoccales population by 21.9% (Tan *et al.* 2011). Due to the electron bifurcation of the heterodisulfide reductase and the difference between reducing potential of cofactor F<sub>420</sub> and ferredoxin (Buckel and Thauer 2013), members of the Methanomassiliicoccales are predicted to have a lower H<sub>2</sub> threshold in comparison to *Methanosphaera* spp. and *Methanosarcinales* spp., which may explain the increased abundance of members of the Methanomassiliicoccales in response to a tannin-rich diet.

The predicted substrates of methylotrophic methanogenesis in Methanomassiliicoccales include methanol, mono-, di-, tri-, methylamines and dimethyl-sulfide (Borrel *et al.* 2014). Previously, methanol was considered to be a methyl substrate common between members of the Methanomassiliicoccales (Lang *et al.* 2015), however, in the analysis reported here, the

comparison to BRNA1 ruled out methanol as a conserved substrate, as the methanol:corrinoid methyltransferase *mtaB* and the corresponding *mtaC* necessary for methanol utilisation are not present in the genome of BRNA1. Only the genes involved in methylamine utilisation are conserved across all of the Methanomassiliicoccales genomes analysed, except RumEn M2 which the incomplete genome lacks known methyl-substrate utilising methyltransferase. All of the ovine rumen strains analysed and two bovine rumen strains are predicted to utilise mono-, di- and tri-methylamine as a substrate, and presumably this reflects the availability of methylamines from plant derived choline and betaine within the rumen (Neill *et al.* 1978; Mitchell *et al.* 1979). Choline is utilised by the rumen protozoan, *Entodinium caudatum*, which incorporates it into phosphatidylcholine, but does not use it for trimethylamine production (Bygrave and Dawson 1976). A choline-utilising gene cluster has been identified in the genera *Desulfovibrio*, *Clostridia*, *Streptococcus*, *Klebsiella*, and *Proteus* (Craciun and Balskus 2012), all of which can be found in the rumen (Huisinigh *et al.* 1974; Widyastuti *et al.* 1992; Ghali *et al.* 2004; Zadoks *et al.* 2011). The genera *Streptococcus*, *Klebsiella* and *Proteus* have also been found inside rumen protozoa, *Entodinium caudatum* (White 1969). Betaine and choline have been shown to be utilised by *Clostridium sporogenes* – *Methanosarcina barkeri* co-cultures (Hippe *et al.* 1979; Naumann *et al.* 1983). Some sulphate utilising marine methanogens, such as *Methanococci* spp., can utilise betaine and choline (L'Haridon *et al.* 2014; Watkins *et al.* 2014). There is no genomic evidence to suggest that members of the Methanomassiliicoccales can use betaine and choline directly as substrates to produce methylamines. Therefore, a close association between betaine and choline degrading bacteria and members of the Methanomassiliicoccales may be advantageous. A member of the Methanomassiliicoccales has been found in co-culture with the anaerobic rumen fungus *Piromyces* sp. (Jin *et al.* 2014). The ovine rumen strains ISO4-H5, ISO4-G1 and ISO4-G11 were enriched with a strain of *Succinivibrio dextrinosolvens* (Jeyanathan 2010) however, previous studies have shown that this species does not utilise betaine (Gomez-Alarcon *et al.* 1982). Nevertheless, a strong association was found between *Succinivibrionaceae* and *Methanomassiliicoccaceae* in a global census of the rumen microbiome (Henderson *et al.* 2015). This association may be explained, at least partially, by the release of methanol by *Succinivibrio* spp. from the degradation of pectin (Dehority 1969), which may be utilised subsequently by members of the Methanomassiliicoccales as a substrate for methanogenesis (Henderson *et al.* 2015).

Methyl-coenzyme M reductase is a core enzyme of methanogenesis, and the holoenzyme is composed of six subunits as an  $\alpha_2\beta_2\gamma_2$  heterohexamer (Ellefson and Wolfe 1981). The genes

that code for the isozymes Mcr and Mrt display different operon structures within a genome, *mcrBDCGA* and *mrtBDGA* respectively (Friedrich 2005). Based on this observation, all members of the Methanomassiliicoccales genomes analysed encode the Mrt isozyme. In organisms which encode both Mcr and Mrt, the Mrt is predominant during early log phase of growth when H<sub>2</sub> and CO<sub>2</sub> are sufficient, whereas the Mcr predominates when growth is limited by available H<sub>2</sub> and CO<sub>2</sub> which also correlates with Mrt having a higher specific activity than Mcr (Rospert *et al.* 1990; Bonacker *et al.* 1993; Pihl *et al.* 1994; Reeve *et al.* 1997). The Mcr/Mrt are highly conserved among methanogens and the sequence of the  $\alpha$  subunit is often used to infer phylogenetic relationships between methanogens (Luton *et al.* 2002; Friedrich 2005; Chaudhary *et al.* 2011). The phylogenetic analysis of McrA (Figure 4.13) suggests the Methanomassiliicoccales Mrt has diverged significantly from the Mrt and Mcr identified from *Methanobrevibacter* and *Methanosphaera* species. A recent metagenome/metatranscriptome study has classified the Mcr/Mrt operons of ovine rumen methanogens into three groups, and represents but not limited to Methanobacteriales Mcr, Methanobacteriales Mrt and Methanomassiliicoccales Mrt respectively, of which the increased transcription of *mcr/mrtA* gene in Methanobacteriales *mcr* group and the Methanomassiliicoccales *mrt* group has been observed in the high methane emitting sheep (Shi *et al.* 2014). In this thesis, a conserved protein has been identified in the *mrt* operon of all 11 genomes, located immediately downstream of the *mrtA* gene. The gene is 91 to 96 bp in size, with conservation in amino acid identity ranging from 31.9% between 1R26 and B10 to 72.8% between ISO4-H5 and ISO4-G11, suggesting an involvement in the Mrt system. This gene will require structural studies to elucidate whether it is a new subunit of the Mrt complex or a protein somehow associated with Mrt (Figure 4.15).

At the central active site of Mrt is a highly reduced tetrapyrrole, coenzyme F<sub>430</sub> (Ermler *et al.* 1997). Members of the Methanomassiliicoccales produce tetrapyrrole by using the glutamyl-tRNA via uroporphyrinogen-III intermediate by enzymes encoded by the *hemABCDL* and *corA* genes (Phillips *et al.* 2003; Lang *et al.* 2015). These genes are highly conserved between eight of the Methanomassiliicoccales genomes analysed, while the ISO4-G11 draft genome only possesses a *hemC* homologue, and all of the genes involved in tetrapyrrole biosynthesis are absent from the ISO4-G1 and RumEn M2 genomes. In *Pseudomonas fluorescens* and *Escherichia coli*, a mutation in the *hemA* gene renders these organisms auxotrophic to the tetrapyrrole biosynthesis intermediate aminolevulinate (Avisar and Beale 1989; Baysse *et al.* 2001). The absence of *hemA* suggests that ISO4-G1, ISO4-G11 and RumEn M2 are incapable



of producing tetrapyrrole, and therefore are likely to require an external source of F<sub>430</sub> or tetrapyrrole intermediates to survive.

CoM is an important co-factor required for the methanogenesis pathway as all methyl groups are transferred to CoM before being reduced to methane (Thauer *et al.* 2008). The *comABCDE* genes are required for CoM biosynthesis (White 1985; Graupner *et al.* 2000; Graupner *et al.* 2000). All the Methanomassiliicoccales genomes sourced from the rumen and the MpT1 genome are predicted to lack *comABCDE*, which suggests they are incapable of CoM biosynthesis, and require an external supply of CoM to survive. The inability to produce CoM has also been observed in *Methanobrevibacter ruminantium* M1<sup>T</sup> (Leahy *et al.* 2010), which is predicted to be able to survive by uptake of external CoM from the rumen environment. Rumen fluid contains 10<sup>7</sup> to 10<sup>9</sup> archaea cells per ml (Hungate 1966), and the rumen archaeal community is composed of almost exclusively of methanogens (Janssen and Kirs 2008). It has been shown that *Mbb. ruminantium* M1<sup>T</sup> requires rumen fluid for growth, and that the supplementation of 3 ng/ml of CoM in the absence of rumen fluid achieves half the maximal growth, indicating that rumen fluid can act as a source of CoM (Taylor *et al.* 1974). A further study has observed the uptake of CoM in *Mbb. ruminantium* M1<sup>T</sup> using labelled H<sup>35</sup>S-CoM (Balch and Wolfe 1979). The availability of CoM in the rumen environment is likely to allow the seven rumen strains of Methanomassiliicoccales lacking CoM biosynthesis functions to survive within the rumen. Similarly, within the termite gut, CoM is available for uptake by organisms such as MpT1 (Lang *et al.* 2015). The differences between the environment in the termite gut (Brune 1998) and the rumen (Hungate 1966) suggests the loss of CoM biosynthesis genes occurred independently. Loss of function can confer adaptive advantage in *E. coli* (Hottes *et al.* 2013). Pairwise competition experiments in *E. coli* and *Acinetobacter baylyi* has shown, in a nutrient-rich environment, evolution favors the loss of biosynthetic genes and the uptake of metabolites that can circumvent the loss of biosynthetic functions (D'Souza *et al.* 2014). Similarly, the loss of the CoM biosynthesis pathway may also confer an adaptive advantage to members of the Methanomassiliicoccales, and enhance their survival in the rumen and termite gastrointestinal environments.

Members of the Methanomassiliicoccales isolated from human faeces have the ability to produce CoM based on their genomic sequences. Methanogens account for approximately 10% of anaerobes in the human colon (Eckburg *et al.* 2005), of which 80% to 100% are *Methanobrevibacter smithii* (Matarazzo *et al.* 2012), which is capable of CoM biosynthesis (Samuel *et al.* 2007). Using the same rationale proposed by Lang *et al.* (2015) for MpT1,

conceivably CoM can be expected in the human gut, and indeed CoM is listed in the Human Metabolome Database as being present (<http://www.hmdb.ca/metabocard> for Mesna (CoM) HMDB03745). However, the three genomes of Methanomassiliicoccales isolated from the human gut have not lost their genes encoding CoM production. Perhaps the difference between the hosts' digestive systems could help explain the difference in CoM biosynthesis capability. Although the members of the Methanomassiliicoccales analysed in this study were isolated from human faeces, it is conceivable that the human colon also harbours members of the Methanomassiliicoccales. Indeed, members of Methanomassiliicoccales (EF628038, EF369488) has been detected in the human colon by *mcrA* gene based diversity analysis (Scanlan *et al.* 2008), although the *mcrA* genes were previously recognized as members of *Methanosarcinales*, subsequent phylogeny studies has verified them to be members of Methanomassiliicoccales (Paul *et al.* 2012; Borrel *et al.* 2013). The retention time of digesta is similar between ovine rumen and the human large bowel (Poppi *et al.* 1981; Van Soest 1984), however, the motion of digesta differs between the human colon and rumen (Hertz and Newton 1913; Phillipson 1939). The sheep rumen has a very regular cycle of contraction as well as regurgitation for rumination (Phillipson 1939), whereas the human colon is completely inactive most of the day with three to four rapid movements lasting only a few seconds (Hertz and Newton 1913). The movement of rumen digesta allows even distribution of CoM throughout the rumen, whereas the inactivity of the human colon may make environmental CoM difficult to diffuse and therefore difficult to access. However, CoM is a very small cofactor of 141 g/mol (Taylor and Wolfe 1974), which suggests the diffusion is only limited to some degree, and there may be a more fitting explanation. The cross-feeding of cofactors has been observed between *Treponema primitia*, *Serratia grimesii*, *Lactococcus lactis* and *Treponema azotonutricium* in the termite gut, where *S. grimesii* and *L. lactis* supplies folate to *T. primitia* (Graber and Breznak 2005), and *T. primitia* supplies biotin, pyridoxal phosphate and coenzyme A to *T. primitia* (Rosenthal *et al.* 2011). Similar cross-feeding may occur between the CoM auxotrophic MpT1 and CoM producing methanogens in the termite gut. The methanogens within human gastrointestinal environment is predominantly *Mbb. smithii* (Matarazzo *et al.* 2012), and it tends to be highly transcriptionally active compare to other species (Franzosa *et al.* 2014). Although *Mbb. smithii* produces CoM, members of Methanomassiliicoccales B10, Mx1 and Mx1201 may not be able to consistently and reliably acquire CoM from *Mbb. smithii*, as biosynthetic pathway for small metabolites tend to be underexpressed (Franzosa *et al.* 2014), possibly CoM biosynthesis pathway is also underexpressed and there is no excess CoM to

facilitate cross-feeding, which may explain why the human strains retained the genetic capability for CoM biosynthesis.

The purpose of methanogenesis is to generate membrane potential and produce energy in the form of ATP (Khakh and Burnstock 2009). Methanogens are strictly chemiosmotic, and only use the archaeal  $A_1A_0$ -ATP synthase, which consists of nine different subunits (Muller *et al.* 1999). The conservation of  $A_1A_0$ -ATP synthase genes between methanogens has been exploited as a potential drug target for methane mitigation (Aung *et al.* 2015). The  $A_1A_0$ -ATP synthase of *Mbb. ruminantium* M1 uses a  $Na^+$  gradient for energy generation, but a  $H^+$  gradient has been found to be used when it is more favourable (McMillan *et al.* 2011). A conserved PET sequence motif in the rotor subunits is associated with  $H^+$  transport in methanogens (Muller *et al.* 1999). A PET motif has been identified in the *ahaK* genes at amino acid 62 – 64 in all the Methanomassiliicoccales genomes analysed, therefore a  $H^+$  gradient may drive energy generation in members of the Methanomassiliicoccales. This is consistent with the  $H^+$ -translocating Fpo-like complex conserved across the Methanomassiliicoccales strains, as well as the lack of the  $Na^+$  gradient-generating Mtr of the hydrogenotrophic methanogenesis pathway (Lienard *et al.* 1996). However, the presence of  $Na^+/H^+$  antiporter and sodium:solute symporters such as sodium/proline symporter PutP and sodium bile acid symporter suggests  $Na^+$  gradient may have a role in transportation of solute in Methanomassiliicoccales.

Members of the Methanosarcinales rely on the Fpo complex and cofactor  $F_{420}H_2$  for energy conservation (Baumer *et al.* 2000) while members of the Methanomassiliicoccales depend on the Fpo-like complex for  $H^+$  gradient generation. The FpoF subunit is involved in interacting with  $F_{420}H_2$  (Welte and Deppenmeier 2011), however, all of the Methanomassiliicoccales genomes, lack the genes required to produce cofactor  $F_{420}$  as well as a gene encoding the FpoF subunit (Borrel *et al.* 2014). FpoO is proposed to be involved in methanophenazine electron transfer (Welte and Deppenmeier 2011). Genes encoding the subunit FpoO are also absent in all the Methanomassiliicoccales genomes, and there are no genes encoding methanophenazine biosynthesis. Members of the Methanomassiliicoccales therefore require an alternative electron carrier in the absence of the cofactor  $F_{420}$ . In *Methanosaeta* spp. which lack the *fpoF* gene, it has been proposed that ferredoxin can act as an alternative electron carrier (Welte and Deppenmeier 2011). Therefore, it is likely that Methanomassiliicoccales ferredoxins are involved in the electron transfer between Hdr, Mvh and the Fpo-like complex. The GIT2 Methanomassiliicoccales RumEn M1, B10 and Mx1 also possess the energy converting-

hydrogenase complex (*echABCDEF*) that could act as an alternative H<sup>+</sup> transporter, however, this may have an anaplerotic role to replenish intermediates (Lie *et al.* 2012).

The main source of reducing potential for methane formation in the Methanomassiliicoccales is H<sub>2</sub>. Ethanol has been known to provide additional reducing potential for methanogenesis (Berk and Thauer 1997; Leahy *et al.* 2010). The NADP-dependent alcohol dehydrogenase is coupled to the F<sub>420</sub>-dependent NADP reductase and a F<sub>420</sub> dependent methylene-H<sub>4</sub>MPT dehydrogenase to allow ethanol to be utilised (Berk and Thauer 1997). A type IV iron containing alcohol dehydrogenase (*adh*) is predicted in the genomes of ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, 1R26 and Mx1201. However, the cofactor F<sub>420</sub> and the F<sub>420</sub> dependent NADP reductase, and F<sub>420</sub> dependent methylene-H<sub>4</sub>MPT dehydrogenase involved in supplying additional reducing potential to methanogenesis are absent in all Methanomassiliicoccales genomes analysed, which suggests the utilisation of ethanol is not linked to methanogenesis. Ethanol may be utilised as an alternative carbon source, but further work is required to define such a role.

The genomes of BRNA1, RumEn M1, RumEn M2, 1R26 and Mx1201 are predicted to encode *ldhA* genes. The Ldh enzyme converts lactate to pyruvate (Furukawa *et al.* 2014), thereby, lactate may serve as a potential carbon source. However, lactate is unlikely to be an alternative carbon source for methanogens (Mackie and Heath 1979), as within the rumen environment, it is expected that methanogens would be unable to outcompete rumen bacteria in the fermentation of lactate. The *Desulfovibrio vulgaris* was only able to catabolise ethanol and lactate in presence of *Methanobacterium bryantii* MoH, as the free energy change of the catabolic reaction is more negative at low H<sub>2</sub> partial pressure, and H<sub>2</sub> utilising *M. bryantii* made this reaction favourable (Bryant *et al.* 1977). Conversely, lactate has been proposed as an electron donor to the Hdr system of methanogenesis in members of the methanogenic order Bathyarchaeota (Evans *et al.* 2015). A lactate dehydrogenase-like FAD-containing dehydrogenase (*glcD*) gene is co-located with *hdrD* genes in the genome sequence of BA1, a member of Bathyarchaeota. The iron sulphur cluster binding motifs of the *hdrD* gene may allow heterodisulfide to utilise lactate as a source of electrons (Evans *et al.* 2015). A *glcD* gene homologue has been found in the Methanomassiliicoccales genomes analysed (AR505\_263, ISO4G1\_0340, ISO4G11\_0756, TALC\_01288, MMALV\_13780, MMINT\_03090, WP\_019177724.1, Mpt1\_c09060) that are not colocalised with the *hdrD* gene, experimental validation is required to confirm whether lactate can be utilised by members of Methanomassiliicoccales.

There are 17 protein-coding genes hypothesized to be marker genes for methanogenic archaea, and they have been termed methanogenesis marker proteins (MMP) (Gao and Gupta 2007). The Methanomassiliicoccales genomes analysed showed conservation of the MMPs as expected, however, they all lack MMPs 9, 10, 12 and 14. Furthermore, BRNA1, 1R26 and Mx1201 lack MMP16, RumEn M2, 1R26 and B10 lacks the MMP2, RumEn M2 lacks the MMP13 and B10 lacks the MMP 4. It is clear from this analysis that not all 17 MMPs can be considered markers of all the methanogenic archaea as previously thought (Gao and Gupta 2007).

Based on the genome sequences of members of the Methanomassiliicoccales, central carbon metabolism is predicted to involve acetate acquisition as the primary carbon source, and use of an incomplete TCA cycle followed by complete gluconeogenesis and pentose phosphate pathways. The acetyl-CoA synthetase gene is conserved in all genomes analysed, therefore acetate is predicted as the common carbon source. The RumEn M1, RumEn M2, Mx1 and B10 genomes contain genes encoding methylene tetrahydrofolate reductase (MetVF), methyltetrahydrofolate:corrinoid methyltransferase (AcsE), carbon monoxide dehydrogenase (AcsA) and acetyl-CoA decarbonylase (CdhDE), which suggests Mx1 and B10 may be able to produce acetyl-CoA from formate and CO<sub>2</sub>. MpT1 may not be able to convert CO<sub>2</sub> to CO, as the primary function of AcsB is in electron transfer (Morton *et al.* 1991), and AcsB homologues can easily be confused with similar electron transfer enzymes. As formate is utilised as an electron donor for hydrogenotrophic methanogens in the rumen (Liu and Whitman 2008), the rumen members of the Methanomassiliicoccales may have evolved to not utilise formate and thus avoid competition with hydrogenotrophic methanogens.

Methanogens are known to have an incomplete TCA cycle (Ekiel *et al.* 1985; Sprott *et al.* 1993), and the Methanomassiliicoccales appear no different as they are predicted to have various genes absent from the TCA pathway (Figure 4.19, Table A.4.16). Despite this, the TCA cycle is still required for gluconeogenesis and nitrogen uptake, in addition to providing intermediates for amino acid biosynthesis. The gluconeogenesis pathway is highly conserved amongst the members of Methanomassiliicoccales, only MpT1 appears to be missing the *pgi* gene encoding the phosphoglucose/phosphomannose isomerase, MpT1 is predicted to utilise an alternative, but unknown, enzyme instead of Pgi (Lang *et al.* 2015). The *pgi* gene exists as a pseudogene in the RumEn M2 and 1R26 genomes due to an interrupted coding sequence, as both genomes are incomplete, this may be caused by sequencing error or a frameshift. The analyses in this thesis also predicts that members of the Methanomassiliicoccales carry out

autotrophic carbon fixation, i.e. the conversion of CO<sub>2</sub> to organic carbon (Berg *et al.* 2010). Ribulose 1,5-bisphosphate carboxylase (RubisCO, RbcL) is the key enzyme in CO<sub>2</sub> fixation and functional type III RubisCO enzymes have been discovered in *Methanococcus jannaschii* and *Pyrococcus kodakaraensis* (Ezaki *et al.* 1999; Watson *et al.* 1999). The archaeal type III RubisCO has been found to function alongside AMP phosphorylase (DeoA) and ribose-1,5-bisphosphate isomerase (E2b2) in AMP metabolism (Sato *et al.* 2007). The RubisCO operon *rbcL*, *e2b2*, *deoA* is found to be conserved across all the Methanomassiliicoccales genomes, and suggests they are capable of fixing CO<sub>2</sub> while recycling AMP. In addition to incorporating carbon fixation to purine recycling, another enzyme in the purine biosynthesis pathway contributes to conservation of cell energy. The genomes of the Methanomassiliicoccales analysed are predicted to use phosphoribosylaminoimidazole carboxylase (Ade2) instead of *N*<sup>5</sup>-carboxyaminoimidazole ribonucleotide synthetase (PurK) to produce NCAIR (*N*<sup>5</sup>-carboxyaminoimidazole ribonucleotide) (Borrel *et al.* 2014). This reduces the ATP required for purine biosynthesis and preserves energy. However, the decrease in energy expenditure may not necessarily correlate with an increase in biomass of the Methanomassiliicoccales, and it remains unknown whether the additional energy would be utilised on growth, maintenance or be dissipated (Russell and Cook 1995).

An analysis of the amino acid biosynthesis pathways in genomes of the Methanomassiliicoccales revealed that several may not be able to make certain amino acids (Table 4.10, Table A.4.17). The Methanomassiliicoccales represent a new order of methanogens and its genes are expected to be different to other methanogens. Therefore it is possible that in instances where a single gene is missing from a pathway, a new gene of unknown function may fulfill the role of the missing gene. In instances where an entire biosynthetic pathway is absent, such as the tryptophan biosynthesis pathway predicted in the genomes of ISO4-G1 and RumEn M2, then it is likely that such organisms require an external source of tryptophan to survive. The Mx1, RumEn M1, RumEn M2 and ISO4-G1 genomes are missing the genes required for a complete tryptophan biosynthetic pathway. The absence of the genes required to produce tryptophan is present in two different genomes of the Methanomassiliicoccales. It has been established that particular traits often evolve repeatedly when the populations are exposed to similar ecological conditions (Schluter *et al.* 2004), which suggests the loss of tryptophan biosynthesis may have evolved in parallel in ISO4-G1, RumEn M1, RumEn M2 of rumen and Mx1 of human gut due to the similarity of the gastrointestinal environment. This is possibly due to the high cost of tryptophan biosynthesis, which costs 78

moles of ATP per mole of tryptophan (Bender 2012). Tryptophan is a scarce but important amino acid in many organisms, as tryptophan confers a strong structural influence by its affinity to aromatic and branched-chain amino acids, repelling negatively charged amino acids as well as stabilizes  $\beta$ -hairpin structures by Trp-Trp binding (Santiveri and Jimenez 2010). Bovine rumen fluid contains 36  $\mu\text{g/ml}$  of tryptophan (Candlish *et al.* 1970). Tryptophan is energy intensive to make, so if cells can uptake tryptophan from the environment, the availability of external tryptophan permitted the removal of unrequired pathway, which conserves energy (Zamenhof and Eichhorn 1967). However, energy is not the only contributing factor to the loss of genes in evolution (Dykhuizen 1978), further investigation is required to clarify the evolutionary driving force that removed tryptophan biosynthesis pathway. The auxotrophic *Methanomassiliicoccales* would likely import the deficient amino acids from the environment via an amino acid permease. An amino acid permease gene is predicted in all genomes analysed except RumEn M1, RumEn M2 and MpT1 (Table A.4.17). All the genomes analysed are also predicted to possess a sodium/proline symporter *putP* gene. All genomes are predicted to be capable of proline biosynthesis.

Comparison between *Methanomassiliicoccales* genomes can reveal conserved genes important for the organisms' growth and survival as well as those genes that might be important for niche adaptation. This study included six completed genomes and five incomplete genomes of *Methanomassiliicoccales*, only the completed genomes were analysed to identify the core genome and environmental conserved gene families. The core genome of the six completed *Methanomassiliicoccales* genomes examined consists of 415 gene families, while the order level pan-genome encompasses 8,767 gene families across 11 genomes. The *Methanomassiliicoccales* strains originated from four environments, the termite gut (1 strain), human faeces (3), and the bovine (4) and ovine rumen (3). In this study, the analysis has been focused on those genomes of rumen origin. There are 18 gene families conserved between the three completed genomes of rumen strains, and 41 gene families conserved between the two completed genomes of ovine rumen strains, which might shed light on the genes involved in adaptation to the rumen environment. Each genome also contains its own unique gene families, from 352 gene families in BRNA1 to 1,137 gene families in B10. These unique gene families can offer insight into how these members of the *Methanomassiliicoccales* may establish their own niche in the presence of similar organisms within the same environment.

The core genome of *Methanomassiliicoccales* was classified according to the COG database (Table 4.7). A total of 89 gene families are implicated in energy conservation, which includes

the genes required to utilise monomethylamine for methanogenesis, but not other substrates. This suggests monomethylamine is the conserved substrate between members of Methanomassiliicoccales. In addition, the aforementioned RubisCo operon is also identified within the core genome, which adds weight to the potential importance and functionality of this autotrophic carbon fixation operon.

There are 18 gene families identified as conserved amongst the genomes of rumen origin (Table 4.8), including but not limited to, a DNA polymerase IV, a methylthioadenosine nucleosidase MtnN, a MIP (major intrinsic protein) transporter, a xanthine/uracil permease family protein and a dolichol kinase. DNA polymerase IV is a Y-family DNA polymerase rarely observed in archaea (Ohmori *et al.* 2001). Due to its lack of proof-reading activity, DNA polymerase IV is more error-prone than other DNA polymerases, which provides this enzyme the ability to replicate DNA across damaged bases, also known as translesion replication (Ohmori *et al.* 2001). This particular enzyme was characterised in *Sulfolobus solfataricus*, and found to be slightly more error-prone than *Taq* polymerase (Boudsocq *et al.* 2001), as it is able to bypass all lesions tested, which was thought to be a selective advantage. The DNA polymerase IV may rescue stalling of DNA replication by translesion replication in members of the rumen Methanomassiliicoccales, allowing more efficient replication and thus cell growth. This may be an adaptation to the rumen, where microorganisms are removed continuously *via* dilution and rumen turnover.

The conservation in the cell envelope related gene, dolichol kinase, within the rumen-dwelling Methanomassiliicoccales might be associated with adaptation to their environment. The cell envelope of Methanomassiliicoccales is likely similar to the glycocalyx envelope of Thermoplasma, with *N*-glycosylated mannose residues and glycolipids (Albers and Meyer 2011). Dolichol kinase is involved in *N*-glycan biosynthesis (Bernstein *et al.* 1989) which suggests the rumen members of the Methanomassiliicoccales have evolved a specialized modification to the cell envelope in order to adapt to the rumen environment.

There are 41 gene families conserved between ISO4-H5 and ISO4-G1 genomes of ovine rumen, including a predicted cardiolipin synthase A, whose biosynthesis is described in Chapter 3, Section 3.2.4. The methanogen cell envelope has been found to be SDS- and protease-resistant (Kandler and König 1978) due to pseudomurein and S-layer, however, *M. luminyensis* B10 has been found to be SDS-sensitive (Dridi *et al.* 2012), and neither electron micrographies nor genomic information suggests the presence of pseudomurein, and



observations thus far does not support the presence of a S-layer ultrastructure, which makes the Methanomassiliicoccales envelope vulnerable to physical and osmotic stress. Cardiolipin has been associated with osmotic stress tolerance (Romantsov *et al.* 2009), as has the ability to tightly bind a transmembrane protein complex, which can assist in salinity tolerance in halophiles (Corcelli 2009). However, cardiolipin synthase of different gene families are also found in all three members of Methanomassiliicoccales from human sources, as well as other methanogens (Yoshinaga *et al.* 2012). Four transporters implicated in cation transport are also conserved between the three ovine rumen Methanomassiliicoccales, including a Na/Pi-co-transporter, a Na<sup>+</sup>/H<sup>+</sup> antiporter family protein as well as two small multidrug resistance proteins that share homology to an ammonium compound efflux pump. The rumen environment has approximately 115 mM of Na<sup>+</sup> (Saleem *et al.* 2013) for both cattle and sheep, whereas phosphate concentrations vary between 1.85 mM to 12 mM in sheep (Wadhwa and Care 2002), and 1 mM to 4 mM in cattle (Saleem *et al.* 2013). The conservation in cation transporters and difference in cardiolipin synthase could potentially aid members of the Methanomassiliicoccales to survive in the rumen environment. The lack of a cell wall makes members of Methanomassiliicoccales vulnerable to osmotic stress, and the conservation of the cation transport genes suggests these genes are involved in maintaining the cellular osmotic balance within the rumen environment.

A large number of gene families (551) were identified as unique to ISO4-H5, including a cluster of genes predicted to be involved in exopolysaccharide production (Chapter 3, Section 3.2.4) and 30 adhesin-like proteins. Members of the Methanomassiliicoccales have been found in acid mine drainage biofilms, where they are predicted to rely heavily on exopolysaccharides for adherence (Mendez-Garcia *et al.* 2014). This particular cluster of putative ISO4-H5 exopolysaccharide producing genes and adhesin-like proteins, may play a role in adherence to solid substrates, host cells or other microorganisms within the rumen. Two adhesin-like protein encoding genes (AR505\_1559, AR505\_1560) are predicted to contain Listeria Bacteroides repeat domains (PF09479) and a cohesin domain (PF00963) (Table 3.10), which may be involved in the adherence of ISO4-H5 to hydrogen producing bacteria with cellulosome.

Another set of genes unique to ISO4-H5 were the *dndCD* genes (AR505\_0884, AR505\_0885) required for DNA phosphorothioation (Wang *et al.* 2007), the process in which DNA is modified by replacing the oxygen in the DNA back bone with sulfur, conferring nuclease resistance (Wang *et al.* 2007). The other genes of the *dnd* operon (*dndABE*) were not found in ISO4-H5, however *dndA* is homologous to cysteine desulfurase, which is found in ISO4-H5

(AR505\_1027), and DndB likely stabilizes secondary DNA structures during phosphorothioation, while the function of DndE remains unknown (Chen *et al.* 2010). In addition, a sulphate permease gene *sulP* (AR505\_1686) is present which may encode supply of the sulfur molecule for phosphorothioation. Therefore, it seems possible that ISO4-H5 possesses the genes encoding phosphorothioation, which may explain why all the previous attempts to verify genome assembly via pulse field electrophoresis resulted in a heavily degraded smear. As when the voltage was applied in the gel electrophoresis, the amine group in Tris (trisaminomethane) buffer was activated, a peracid oxidant is generated that carries out nucleolytic attack at the modified DNA sites (Chen *et al.* 2010).

Other noteworthy genes unique to ISO4-H5 include a DNA alkylation repair enzyme AR505\_1309 and a 6-*O*-methylguanine DNA methyltransferase *ogt* (AR505\_0448) gene that repairs alkylating lesions, and a phosphoglycolate phosphatase (AR505\_0604) thought to salvage the phosphoglycolate produced by the RubisCO operon during photorespiration. The presence of this gene is unexpected as ISO4-H5 is an obligate anaerobe, and the RubisCO most likely only catalyse carbon fixation that produces 3-phospho-D-glycerate, and not photorespiration.

The ISO4-G1 genome contains 526 unique gene families (Table A.4.13) and it is noteworthy that it has a high number of predicted ferric ion transporters. In *Methanosarcina barkeri* ferric ions directly inhibited methanogenesis, as ferric ion is preferentially reduced over the methyl-substrates (Bodegom *et al.* 2004). There is no experimental evidence to suggest the high number of transporters predicted to be involved in iron transporter can be correlated to the utilisation of iron in ISO4-G1, but as Fe-S centre containing enzymes occupy important metabolic roles in methanogen, the uptake of iron may benefit ISO4-G1. Additionally, there is a large NRPS identified as unique to the ISO4-G1 genome (Kelly *et al.*, 2016). Archaeal NRPSs were first described in *Methanobrevibacter ruminantium* M1<sup>T</sup> (Leahy *et al.* 2010), and subsequently discovered in other classes of *Methanomicrobia* and *Methanobacteria* (Wang *et al.* 2014) and *Mbb. millerae* SM9 (Kelly *et al.* 2016). NRPSs are responsible for biosynthesis of small peptides, and a number of peptides produced by NRPSs are currently used as antibiotics, including penicillin and bacitracin (Witting and Sussmuth 2011). NRPSs are produced by consecutive condensation of substrates, including amino acids, various fatty acids, hydroxyl acids and nonproteinogenic amino acids (Caboche *et al.* 2008). The NRPSs are composed of distinct modules; an adenylation domain that activates the aminoacids, a peptidyl carrier domain that propagates the peptide chain, and a condensation domain that catalyses the

condensation of the amino acids (Strieker *et al.* 2010). Keeping a gene as large as a NRPS in the genome is energetically expensive therefore it seems likely that the gene provides certain advantages for ISO4-G1 survival within the rumen. However the actual benefit of NRPS within the rumen environment, are not currently known, but they may possibly mediate interactions between other rumen microbes.

There were 465 gene families unique to the ISO4-G11 genome (Table 4.S.10), of which 354 genes were COG classified as poorly characterized without predicted function. There were 14 genes involved in membrane biogenesis, including a cluster of five genes that are predicted to be involved in dTDP-L-rhamnose biosynthesis, six genes predicted to be involved in glycosylation of lipopolysaccharide, and two Mur ligase genes possibly involved in glycopeptide formation. Glycosylated dTDP-L-rhamnose has been found attached to the S-layer glycoprotein in *Haloferax volcanii* (Kaminski and Eichler 2014) and the presence of these genes in ISO4-G11 suggests it may possess a cell envelope with a glycosylated exopolysaccharide. A 3,644 aa adhesin-like protein (ISO4G11\_1719) is homologous to a filamentous hemagglutinin. Filamentous hemagglutinins are utilised by *Bordetella pertussis* for adherence to the host respiratory tract (Kimura *et al.* 1990), which suggests ISO4-G11 may use this adhesin for adherence to the ruminant host.

There were 352 gene families unique to the BRNA1 genome, of which 254 gene families were COG classified as poorly characterized without predicted function. Like ISO4-G11, several unique BRNA1 genes have been predicted to be involved in cell membrane biogenesis. Amongst the unique gene families are two genes homologous to phosphoenolpyruvate phosphomutase *fom1* (TALC\_00083) and phosphoenolpyruvate decarboxylase *fom2* (TALC\_00082). These genes form part of the fosfomycin biosynthesis pathway in *Streptomyces* spp. (Kim *et al.* 2012), and catalyse the proton dependent conversion of phosphoenolpyruvate to phosphonoacetaldehyde (Kim *et al.* 2012). A 2-aminoethylphosphonate aminotransferase homologue (TALC\_00081) clusters with *fom1* and *fom2* genes, which could interconvert phosphonoacetaldehyde to 2-aminoethylphosphonate (Dumora *et al.* 1983). The 2-aminoethylphosphonate has been identified as a constituent of phospholipid in the ciliate protozoa *Tetrahymena pyriformis* (Thompson 1969) and its mitochondria (Jonah and Erwin 1971). In addition, two genes were annotated as glycosyltransferases involved in cell wall biogenesis (TALC\_00068 and TALC\_00105). The BRNA1 gene complement suggests it is capable of producing 2-aminoethylphosphonate, but

whether it can act as a constituent of its cell membrane phospholipid requires further investigation.

#### **4.4 Conclusion**

Members of the order Methanomassiliicoccales have been discovered from a wide range of environments, however only members of the Methanomassiliicoccales isolated from the gastrointestinal environments of animals have been cultured and genome sequenced thus far. The sequencing of four members of the Methanomassiliicoccales isolated from the rumen has provided valuable insights into the lifestyle and adaptation of these methanogens to the rumen. All Methanomassiliicoccales genomes of rumen origin are predicted to have lost the genetic capability to produce the essential coenzyme M cofactor, and rely on uptake of CoM directly from the rumen environment. This streamlining of their genomes likely confers a competitive edge over CoM-producing strains due to savings in energy production and in speed of replication. Furthermore, the RumEn M2, ISO4-G1 and ISO4-G11 genomes are predicted to have lost the capability to produce the cofactors tetrapyrrole and F<sub>430</sub>, and requires external uptake of these cofactors to survive. Again, if these cofactors can be easily obtained from the rumen environment, it could reduce energy expenditure by these organisms and offer them an advantage over their co-factor-producing competitors. The continual dilution and turnover of rumen contents makes growth rate and replication speed key factors in persistence within the rumen. The conservation of DNA polymerase IV in rumen Methanomassiliicoccales genomes may also reflect these selection pressure, such that genome replication accuracy is sacrificed for speed of replication.

The rumen members of the Methanomassiliicoccales are predicted to primarily utilise methylamines, which are likely derived from choline and betaine breakdown in the rumen. Comparison of the completed genomes has revealed that mono-methylamine is the only common substrate for methylotrophic methanogenesis, which was also identified amongst incomplete genomes with exception to RumEn M2. ISO4-H5, ISO4-G11, 1R26, Mx1201 and B10 have evolved to utilise a wide variety of methyl-substrates, while ISO4-G1, BRNA1, RumEn M1, Mx1 and MpT1 has utilise only a small selection of methyl-substrate. The utilisation of methylamines requires pyrrolysine incorporation in the methyltransferase enzymes, and the biosynthesis and incorporation of pyrrolysine is conserved across all genomes analysed of the members of the Methanomassiliicoccales, except the incomplete RumEn M2

genome. However, it appears that members of GIT1 cluster and GIT2 likely regulate the amber-codon read through differently.

Very little is known about cell envelope biosynthesis in the Methanomassiliicoccales. They are thought to lack a rigid S-layer and pseudomurein cell wall, which suggests the cell envelope is vulnerable to osmotic stress and requires proficient ion channel to maintain cellular osmolarity. The presence of cardiolipin and dolichol kinase genes in their genomes suggest the cell envelope is similar to the glycocalyx envelope of members of the Thermoplasma. This envelope is likely flexible and potentially covered with *N*-glycosylated mannose residues, glycolipids and other exopolysaccharides that aid in binding to the host or other beneficial microorganisms.

The 11 Methanomassiliicoccales representatives from four different environments revealed genomic elements that encodes central metabolism in order Methanomassiliicoccales, furthermore, conservation between the rumen members of Methanomassiliicoccales have revealed specialised DNA polymerase, cell envelope and cation transport genes that are important to the adaptation of environmental conditions, and genes predicted to be involved in exopolysaccharide production and adhesin-like proteins that are unique to ISO4-H5, these findings will require physiological and biochemical studies to verify.



## Chapter 5

### Phenotypic characterisation of ISO4-H5 and analysis of its growth under high and low hydrogen levels

#### 5.1 Introduction

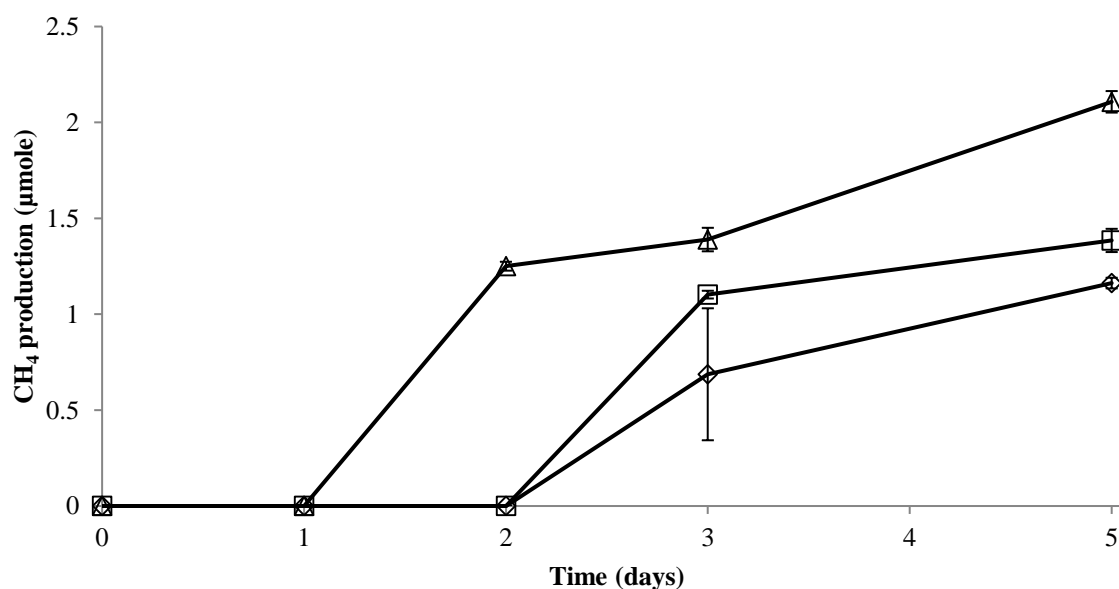
As described in Chapter 3, methanogen cultivation studies have been undertaken in order to understand the types of Methanomassiliicoccales present in the ovine rumen (Jeyanathan 2010). The Methanomassiliicoccales enrichment cultures displayed a common coccoid morphology and lacked autofluorescence at 420 nm in these putative methanogen cells (Jeyanathan 2010). Previous attempts were made to isolate the Methanomassiliicoccales organisms from the ovine rumen enrichment cultures, including 10-fold serial dilutions, application of various antibiotics and heat treatments and although pure cultures of Methanomassiliicoccales were not obtained, the bacterial inhabitants in the enrichment cultures were reduced in both number and diversity (Jeyanathan, 2010). Analysis of the bacterial partner organisms in the enrichments showed a single strain of *Succinivibrio dextrinosolvens* (NR026476, 99% 16S rRNA gene identity) was present (Jeyanathan 2010), suggesting there may be specific interactions between this organism and Methanomassiliicoccales (Jeyanathan 2010). Additionally, a recent global rumen census project indicates a correlation between rumen Succinivibrionaceae with some rumen Methanomassiliicoccales

Recent studies have provided some insights into Methanomassiliicoccales metabolism. Methanomassiliicoccales carries out hydrogen-dependent methylotrophic methanogenesis, utilizing mono-, di-, tri-methylamines, methanol and dimethyl-sulfide (Borrel *et al.* 2014). Despite the understanding gained from genomes of Methanomassiliicoccales of human and termite gut origin, it is important to study the Methanomassiliicoccales from the rumen in order to understand their function in the rumen and to assess their contribution to CH<sub>4</sub> production in this environment. Therefore, studies were carried out on the ISO4-H5 Methanomassiliicoccales enrichment culture with the aim of attempting isolation of the methanogenic archaeon from the enrichment, to enable its growth requirements to be determined and exploring its response to different levels of hydrogen in co-culture experiments with the cellulose-degrading, hydrogen-producing bacterium, *Ruminococcus flavefaciens*.

## 5.2 Results

### 5.2.1 Isolation of ISO4-H5

A *S. dextrinosolvens* strain from the ISO4-H5 enrichment culture was isolated previously by colony picking from BY plates and designated as *S. dextrinosolvens* strain H5 (Cox *et al.*, pers. comm.) and was kindly supplied for this study. During growth studies of the *S. dextrinosolvens* H5 culture, pectin (1% w/v) was accidentally introduced to the ISO4-H5 enrichment culture. The culture was not discarded, but was monitored, and elevated CH<sub>4</sub> production relative to the control culture without added pectin, was observed. Further testing was conducted on supplementation of the ISO4-H5 enrichment cultures with glucose (10mM) and pectin (1% w/v). CH<sub>4</sub> production of ISO4-H5 enrichment cultures were stimulated by both glucose and pectin supplementation, however, pectin supplementation stimulated CH<sub>4</sub> production earlier and to a greater extent at day five ( $p = 5.6\text{E-}05$ ) than glucose ( $p = 1.4\text{E-}02$ ) (Figure 5.1).



**Figure 5.1 Stimulation of CH<sub>4</sub> production in ISO4-H5 enrichment cultures triplicates by glucose or pectin supplementation.** All replicates contain formate (60 mM), methanol (20 mM), acetate (20 mM), CoM (10 μM) and 0.1× vitamin mix. (—◇—) Control, (—□—) glucose (10 mM final concentration) or (—Δ—) apple pectin (1% w/v final concentration) was supplied to ISO4-H5 enrichment cultures, and CH<sub>4</sub> production per ml of headspace was measured daily. Error bars represent +/- standard errors of the means (SEMs). ISO4-H5 growth control is the standard cultivation conditions for the ISO4-H5 enrichment culture previously described (Chapter 2, section 2.2.1.)

From these observations, it was hypothesized that pectin addition indirectly stimulated CH<sub>4</sub> production by enhancing *S. dextrinosolvens* H5 growth, which in turn supplied a growth-limiting metabolite to the methanogenic archaeon in the ISO4-H5 enrichment culture. It was



reasoned that such a growth-limiting metabolite was likely to be soluble and present in the supernatant of *S. dextrinosolvens* cultures grown on pectin. If so, the supernatant itself would allow the growth of the methanogenic archaeon in the absence of *S. dextrinosolvens* H5 cells, therefore *S. dextrinosolvens* H5 cultures were grown for two days in BY medium containing pectin (1% w/v). The bulk of *S. dextrinosolvens* H5 cells were removed by filtering through a 0.22 µm filter to remove cells, thus filter sterilising the preparation. Although previous attempts to remove *S. dextrinosolvens* H5 with antibiotic treatments were unsuccessful, the sensitivity of *S. dextrinosolvens* and the methanogenic archaeon in the ISO4-H5 enrichment culture to antibiotics was determined. The antibiotics were added to the ISO4-H5 enrichment culture, and the culture was also supplemented with glucose (10 mM final concentration) and growth of *S. dextrinosolvens* H5 after overnight incubation was recorded. The incubation was continued for seven days after which CH<sub>4</sub> production from the antibiotic tested culture was measured (Table 5.1).

**Table 5.1 Antibiotic application to enrichment culture**

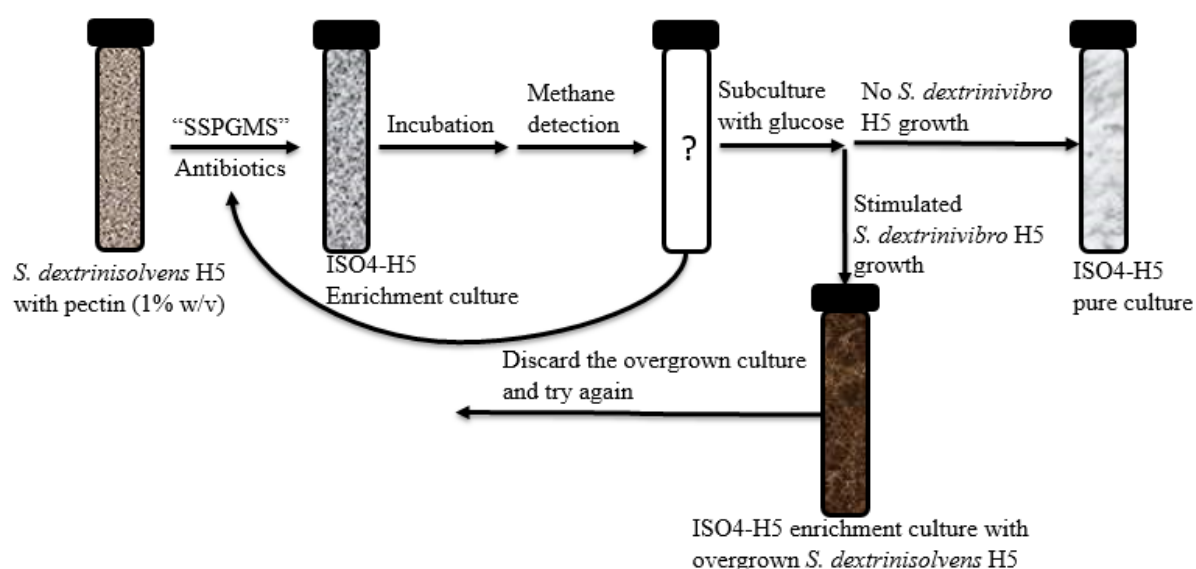
Antibiotic	Final concentration (µg/ml)	<i>S. dextrinosolvens</i> H5 growth overnight	CH <sub>4</sub> production after 7 days
Ampicillin	50	Yes	Yes
	100	Yes	Yes
	250	No	No
	500	No	No
Chloramphenicol*	50	No	No
Erythromycin*	20	Yes	Yes
Gentamycin	20	Yes	No
Kanamycin	50	Yes	Yes
Rifampicin <sup>#</sup>	1.2	Yes	Yes
Spectinomycin	100	Yes	Yes
Streptomycin	100	Yes	Yes
Tetracycline*	10	Yes	Yes
	20	Yes	Yes
	50	Yes	Yes
	100	Yes	Yes

\* dissolved in ethanol. # dissolved in methanol at maximum antibiotic solubility.

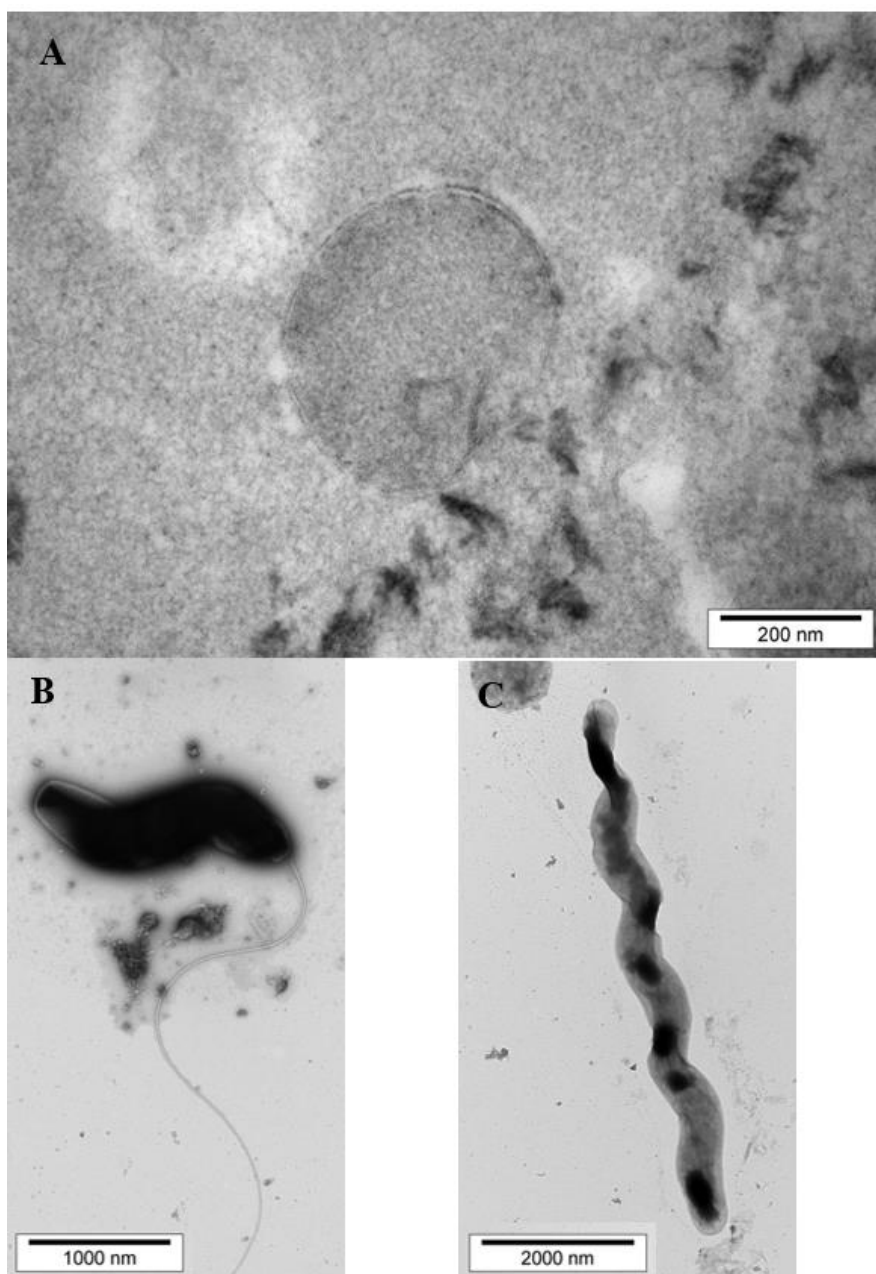
Only chloramphenicol and ampicillin inhibited *S. dextrinosolvens* H5, while only gentamycin inhibited CH<sub>4</sub> production in the ISO4-H5 enrichment cultures. In all cases of *S. dextrinosolvens* H5 inhibition, there was no CH<sub>4</sub> detected after seven days of incubation, indicating that the methanogenic archaeon did not grow, possibly due to direct inhibition by the antibiotics, or possibly because ISO4-H5 did not survive in the absence of *S. dextrinosolvens* H5 growth.

The ampicillin results were particularly encouraging as archaea are insensitive to ampicillin, and suggested a concentration between 100 and 250 µg/ml might be useful for inhibiting *S. dextrinosolvens* growth. The isolation procedure used is outlined in Chapter 2 (section 2.2.3, Figure 5.2.), but briefly it involved supplementing ISO4-H5 enrichments with cell-free, spent

growth medium filtered from pectin-grown *S. dextrinosolvens* H5 cultures which is termed *Succinivibrio* Spent Pectin Growth Medium Supernatant (SSPGMS). In addition to supplementing filter-sterilised SSPGMS and with ampicillin and incubating the cultures until CH<sub>4</sub> was detected in the headspace. The culture was then sub-cultured onto fresh medium containing the same supplements, while a parallel inoculum was introduced to medium containing glucose and incubated overnight to check for the presence of remaining *S. dextrinosolvens*. Initial attempts using 100 µg/ml ampicillin were unsuccessful and *S. dextrinosolvens* persisted in the enrichment culture. The ampicillin concentration was increased to 200 µg/ml, and the isolation procedure was repeated through three cycles of SSPGMS supplementation and ampicillin treatment before the glucose supplemented test cultures remained clear and CH<sub>4</sub> was detected in the culture headspace. The putative pure methanogen culture consisted of coccoid cells that usually do not fluoresce at 420 nm. The putatively pure cultures were screened with PCR amplification using primers specific for both bacterial and archaeal 16S rRNA genes. The bacterial 16S rRNA PCR failed to produce an amplification product, while the archaeal 16S rRNA gene PCR produced a product of the expected size. The PCR product was sequenced and identified as a methanogenic archaeon, affiliated with the order Methanomassiliicoccales, and the culture from which this amplified DNA originated was designated as strain ISO4-H5.



**Figure 5.2 Isolation procedure for ISO4-H5.** SSPGMS: *Succinivibrio* Spent Pectin Growth Medium Supernatant. Antibiotics: 200 µg/ml ampicillin. Glucose: 10 mM. Length of incubation: one week.

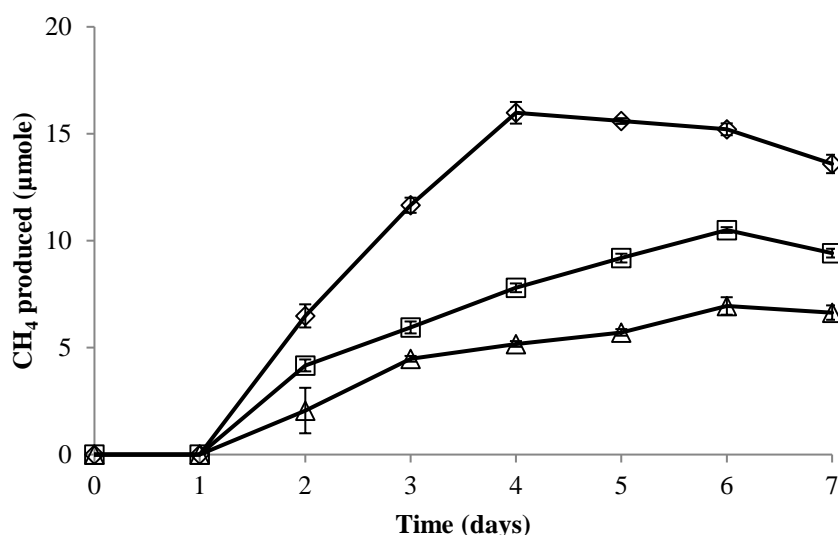


**Figure 5.3 Transmission electron micrograph (TEM) of methanogenic archaeon ISO4-H5 and *S. dextrinosolvens* H5.** A. TEM of negatively stained thin section of the methanogenic archaeon ISO4-H5. B and C. TEM of negatively stained *S. dextrinosolvens*. The samples were prepared as previously described in Chapter 2. Images were captured using a Philips CM10 Transmission Electron Microscope, using an Olympus SIS Morada camera and SIS iTEM software (Germany)

The morphologies of *S. dextrinosolvens* H5 and methanogenic archaeon ISO4-H5 were examined under transmission electron microscopy (TEM) using negative staining of cultures, and ultrathin sectioning of ISO4-H5 was also carried out. The methanogenic archaeon ISO4-H5 has a thin, bi-layer cell membrane with no visible external features (Figure 5.3A). An unidentified structure can be seen within the cell. The negatively stained *S. dextrinosolvens* H5

preparations show it possess a single terminal flagellum (Figure 5.3B), and some cells displayed an elongated spiral morphology (Figure 5.3C). The size of ISO4-H5 was approximately 0.4  $\mu\text{m}$  in diameter, whereas *S. dextrinosolvens* H5 was approximately 0.5  $\mu\text{m}$  in diameter and varied in length from 1.5  $\mu\text{m}$  to over 7.0  $\mu\text{m}$ . Observation of the ISO4-H5 enrichment culture under phase contrast microscopy did not see any evidence of specific physical interaction between the methanogenic archaeon ISO4-H5 and *S. dextrinosolvens* H5.

The growth of the ISO4-H5 pure culture was compared with that of the enrichment culture (Figure 5.4). The enrichment culture produced significantly more  $\text{CH}_4$  than the pure culture, and the SSPGMS stimulated  $\text{CH}_4$  production in ISO4-H5 pure culture, although it is not necessary for pure culture to thrive (Figure 5.4).

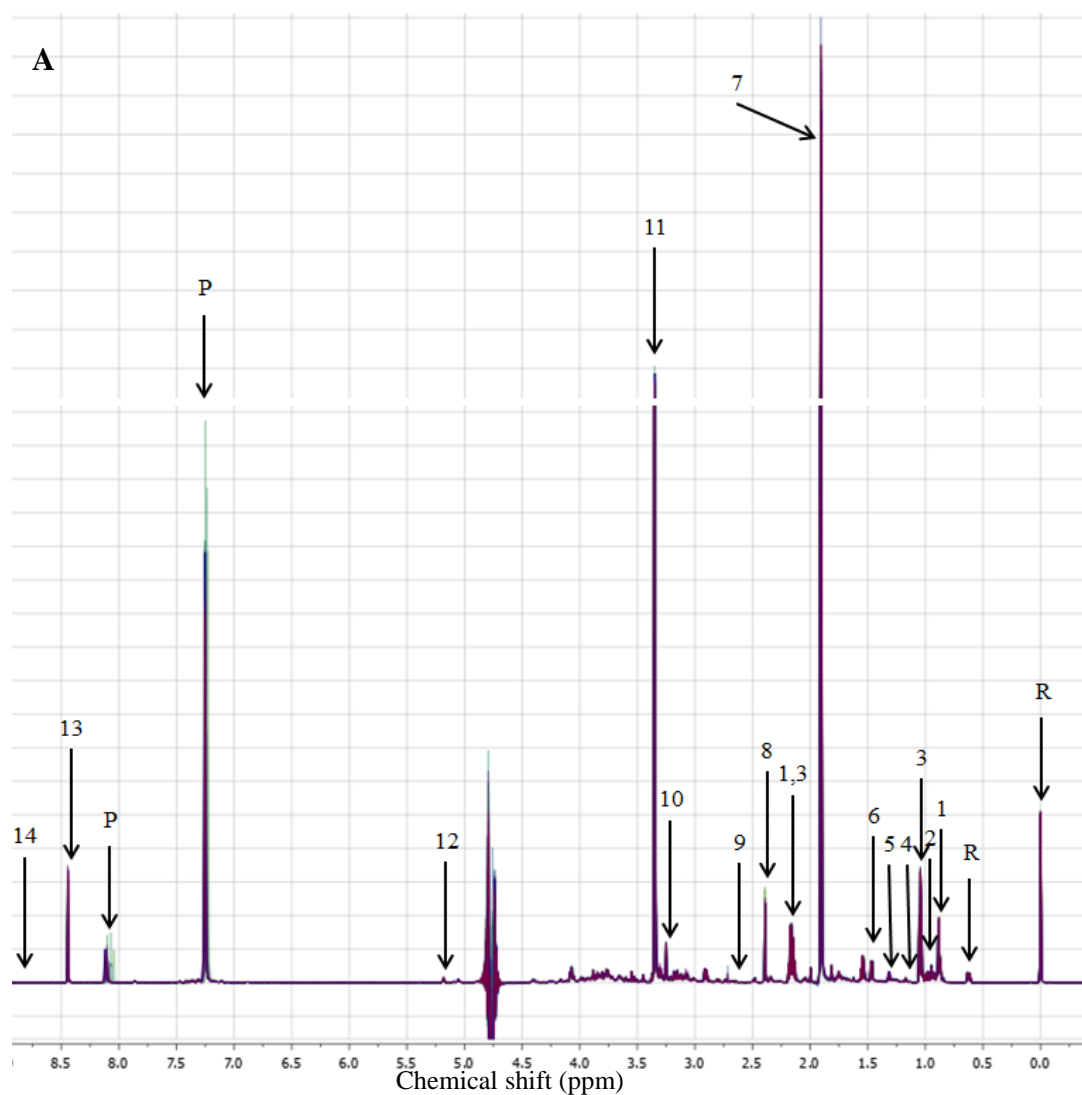


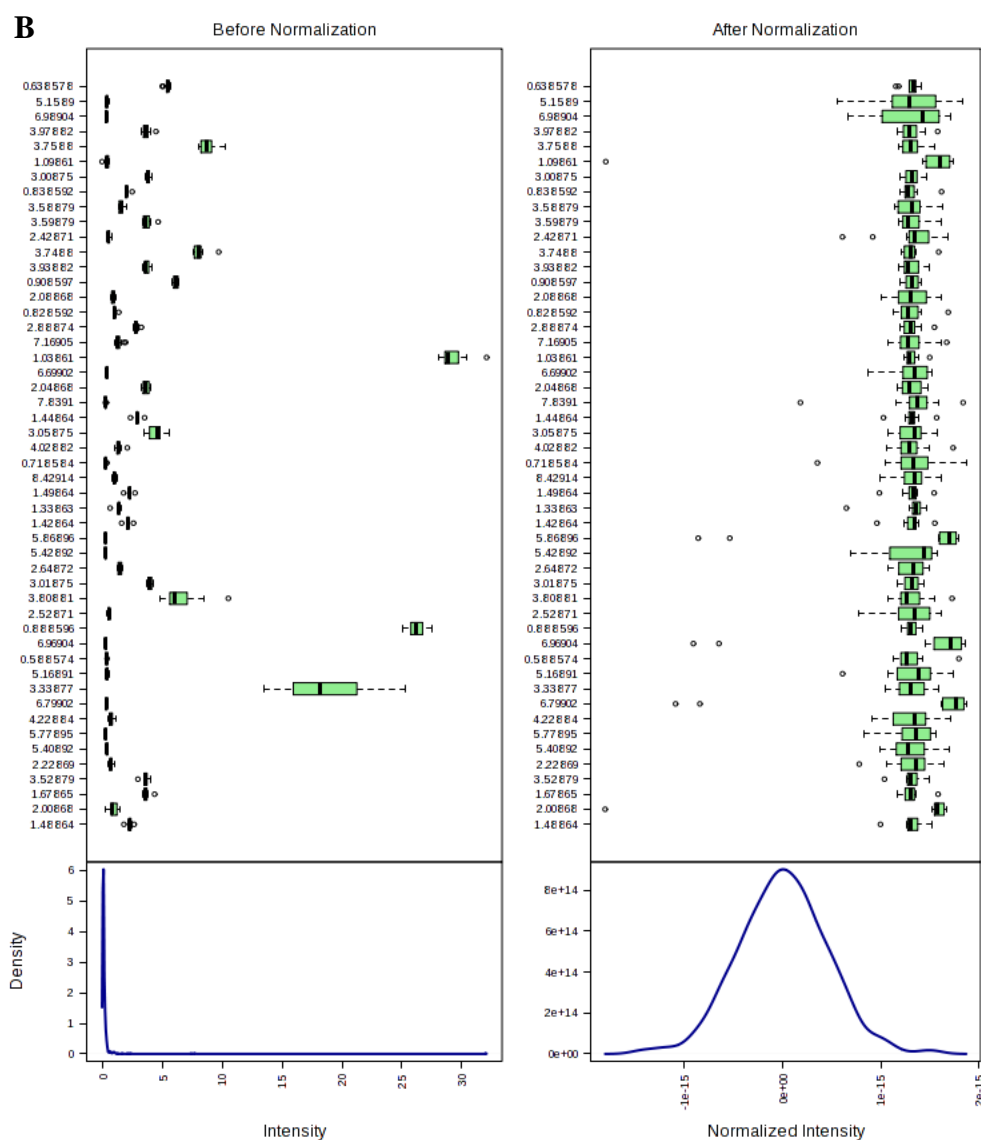
**Figure 5.4 Comparison of growth between pure and enrichment cultures triplicates of ISO4-H5.** All replicates contain formate (60 mM), methanol (20 mM), acetate (20 mM), CoM (10  $\mu\text{M}$ ) and 0.1 $\times$  vitamin mix, (—◇—) ISO4-H5 enrichment culture, (—□—) ISO4-H5 pure culture with 10% (v/v) SSPGMS, (—△—) ISO4-H5 pure culture without SSPGMS.  $\text{CH}_4$  per ml of headspace was measured and expressed as  $\mu\text{mole}$  per ml. Error bars represent  $\pm$  SEMs.

### 5.2.2 NMR analysis of SSPGMS-supplemented ISO4-H5 culture

To identify the components of SSPGMS that stimulated the growth of ISO4-H5 and allowed its isolation, proton nuclear magnetic resonance ( $^1\text{H}$  NMR) was used. The purified ISO4-H5 was cultured in medium with SSPGMS supplementation, and the difference in the  $^1\text{H}$  NMR spectra of the medium before and after cultivation was compared (Figure 5.5). Uninoculated

controls (medium with SSPGMS added, and incubated for the same period of time) were also analysed to rule out changes in medium composition over the duration of the incubation. The concentrations of detected metabolites estimated from the spectra are listed in Table 5.2.





**Figure 5.5 NMR analysis of medium composition before and after growth of ISO4-H5.** **A.** Overlay of all spectra collected from 24 samples (6 inoculated replicates pre and post incubation, 6 uninoculated replicates pre and post incubation). R, internal standard 4,4-dimethyl-4-silapentaine-1-sulfonic acid (DSS); 1, butyrate; 2, branched-chain amino acids; 3, propionate; 4, ethanol; 5, lactate; 6, alanine; 7, acetate; 8, succinate; 9, dimethylamine; 10, betaine; 11, methanol; 12, trehalose; 13, formate; 14, nicotinamide; P, pH indicator – imidazole. **B.** Box plots and kernel density plots of binned spectra of sample replicates before and after normalisation by generalized log transformation (glog 2) and Pareto scaling (mean-centered and divided by the square root of standard deviation of each variable). The chemical shift of spectra bins does not correlate with spectra peaks displayed in A. Intensity of resonance signals and density of spectra bins are displayed. The density plot shown in the bottom panel summarizes the distribution data from all samples. The boxplots show at most 50 features due to space limit.

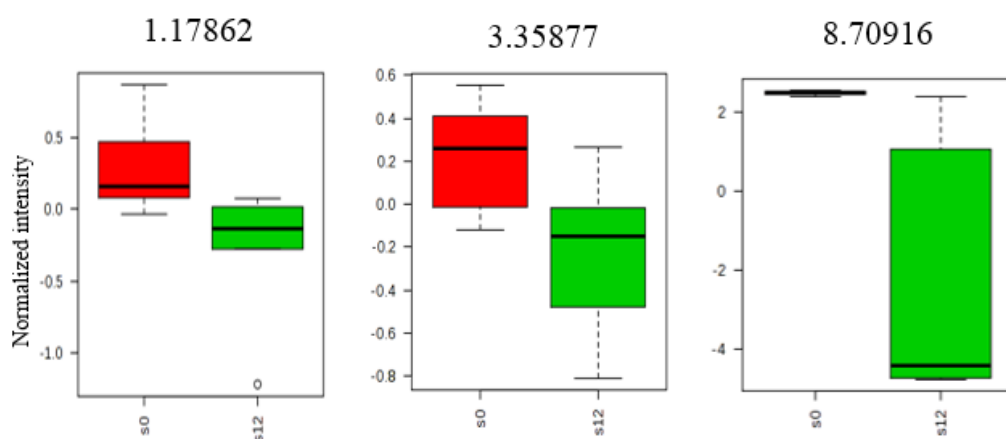


**Table 5.2 Metabolite concentrations estimated by <sup>1</sup>H NMR spectroscopy\***

Compounds	Before growth		After growth	
	mM*	SD	mM*	SD
Butyrate	1.190	0.035	1.123	0.116
2-hydroxyvalerate	0.134	0.009	0.139	0.007
Leucine	0.245	0.014	0.239	0.020
Isoleucine	0.170	0.019	0.172	0.011
Valine	0.192	0.007	0.193	0.022
Propionate	2.440	0.048	2.396	0.125
Ethanol	0.131	0.023	0.100	0.010
Lactate	0.130	0.010	0.135	0.017
Alanine	0.300	0.013	0.300	0.010
Acetate	25.565	1.503	24.677	1.422
Acetamide	0.187	0.015	0.182	0.051
Succinate	0.458	0.059	0.452	0.053
Methionine	0.038	0.008	0.034	0.008
Aspartate	0.164	0.016	0.216	0.022
Dimethylamine	0.023	0.017	0.020	0.012
Phenylalanine	0.102	0.006	0.102	0.025
Betaine	0.116	0.006	0.116	0.003
Methanol	17.886	0.899	15.282	0.605
Trehalose	0.126	0.006	0.127	0.004
Tyrosine	0.043	0.005	0.049	0.003
Formate	4.432	0.413	4.264	0.473

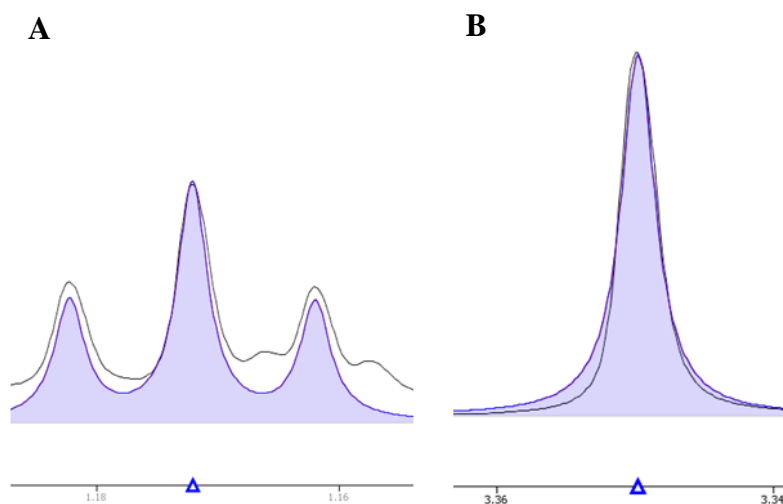
\*: numbers indicate mean concentrations from six replicates estimated from 1D <sup>1</sup>H NMR using DSS as reference

The spectra were binned and normalized as described previously (Chapter 2, Section 2.2.27) and analysed statistically by Principal Component analysis and *T*-test (Figure 5.6). The paired binned spectra that displayed statistical significance of  $P \leq 0.05$  are listed in Appendix Tables A.5.1, and A.5.2. The spectra bins that displayed significant difference before and after growth in both control and sample group were disregarded. The spectra bins 1.18 and 8.71 were significantly reduced post-growth in paired comparisons where each replicate were compared before and post-incubation individually, while spectrum bin 3.35 was significantly reduced post-growth in an unpaired comparison where the before and post-incubation spectra were compared between all replicates (Figure 5.7).

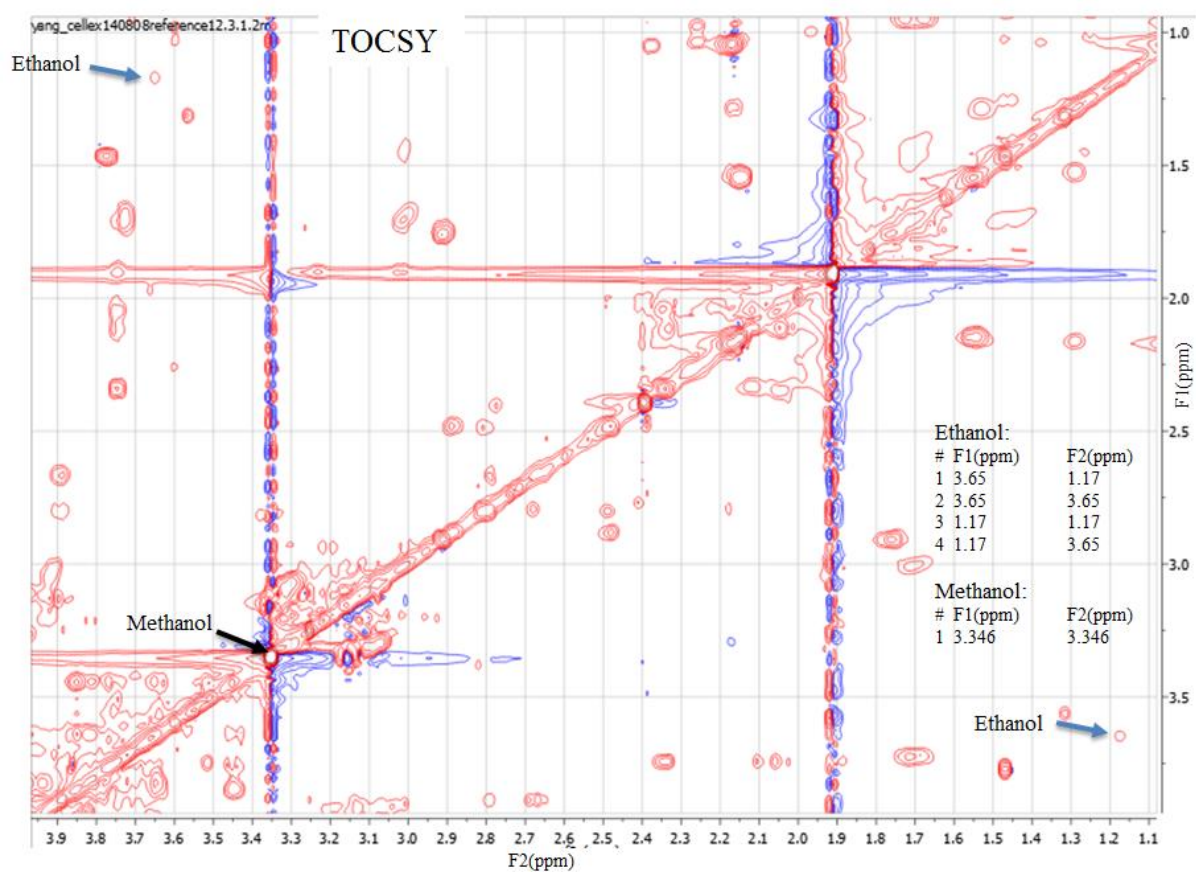


**Figure 5.7 Spectra bins that were reduced significantly post-growth.** s0: samples before incubation, s12: samples post-incubation. Spectra resonances were normalized by log transformation and Pareto scaling.

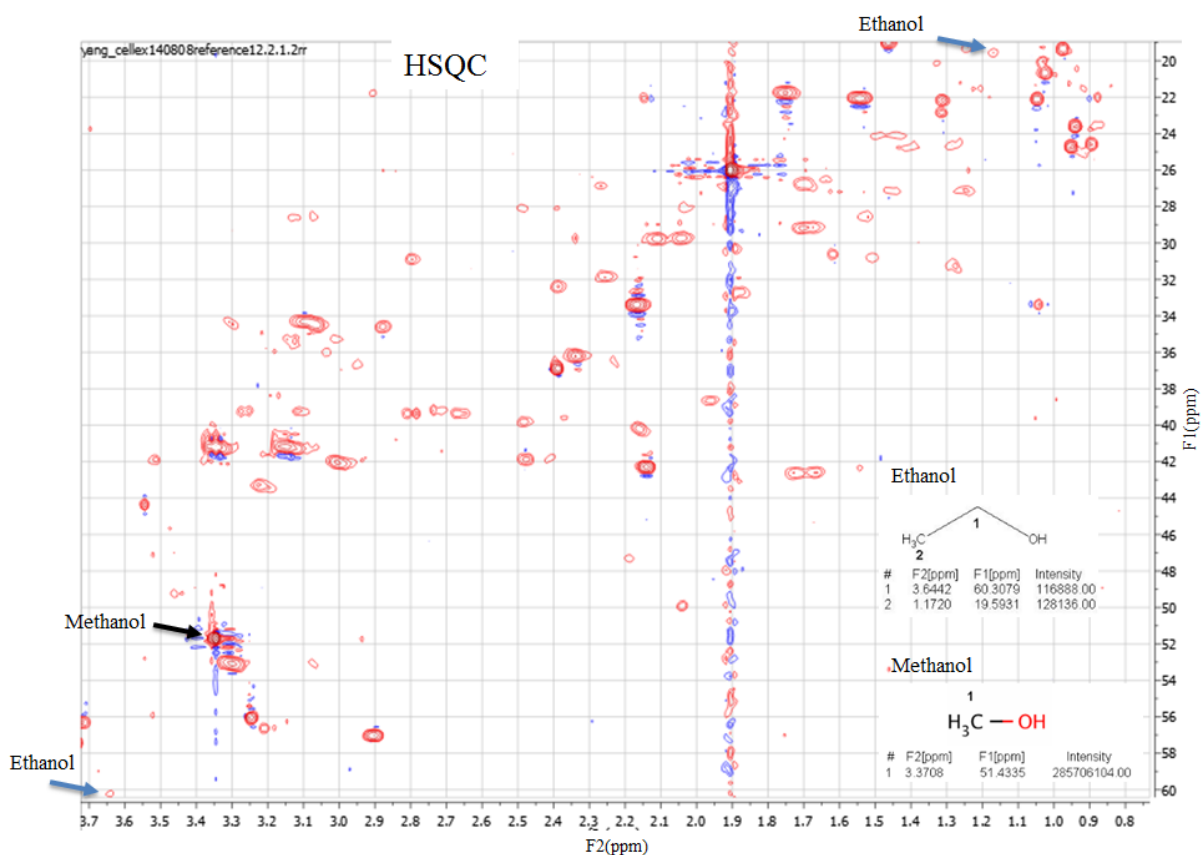




**Figure 5.8 One-dimensional peak alignments.** A. Chemical shift 1.18 aligned with standard peak of ethanol. B. Chemical shift 3.35 aligned with standard peak of methanol.



**Figure 5.9 TOCSY of pooled samples.** The correlation between protons (F1) and protons (F2) from samples with ethanol and methanol are displayed as positive (red) and negative (blue) peaks. The chemical shift values for two dimensional spectrum (F1 and F2) corresponding to ethanol and methanol are displayed. The sample spectrum associated with ethanol is labelled with blue arrows, the sample spectrum associated with methanol is labelled with black arrow.

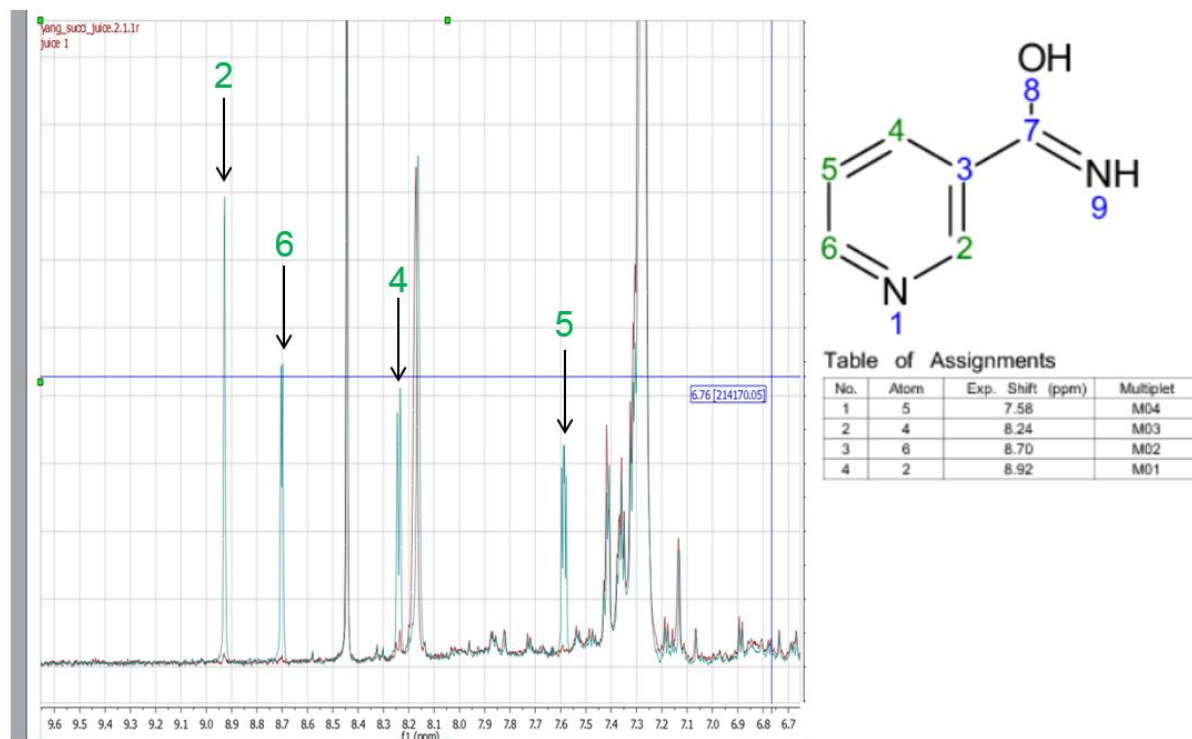


**Figure 5.10 HSQC of pooled samples.** The correlation between protons (F2) and carbons (F1) for ethanol and methanol are displayed as positive (red) and negative (blue) peaks. The chemical shift values for two dimensional spectrum (F1 and F2) corresponding to ethanol and methanol are displayed. The spectrum associated with ethanol is labelled with blue arrows, the spectrum associated with methanol is labelled with black arrow.

Identification of the compounds associated with each spectrum bin was carried out by one-dimensional peak alignment, two dimensional Heteronuclear Single Quantum Coherence Spectroscopy (HSQC), and Total Correlated Spectroscopy (TOCSY) mapping with reference compounds. A compound with a chemical shift of 1.18 was identified as ethanol by one-dimensional peak alignment, and a compound with chemical shift of 3.35 was identified as methanol by one-dimensional peak alignment (Figure 5.8). Both the ethanol and methanol were also verified in pooled sample TOCSY and HSQC (Figure 5.9, 5.10).

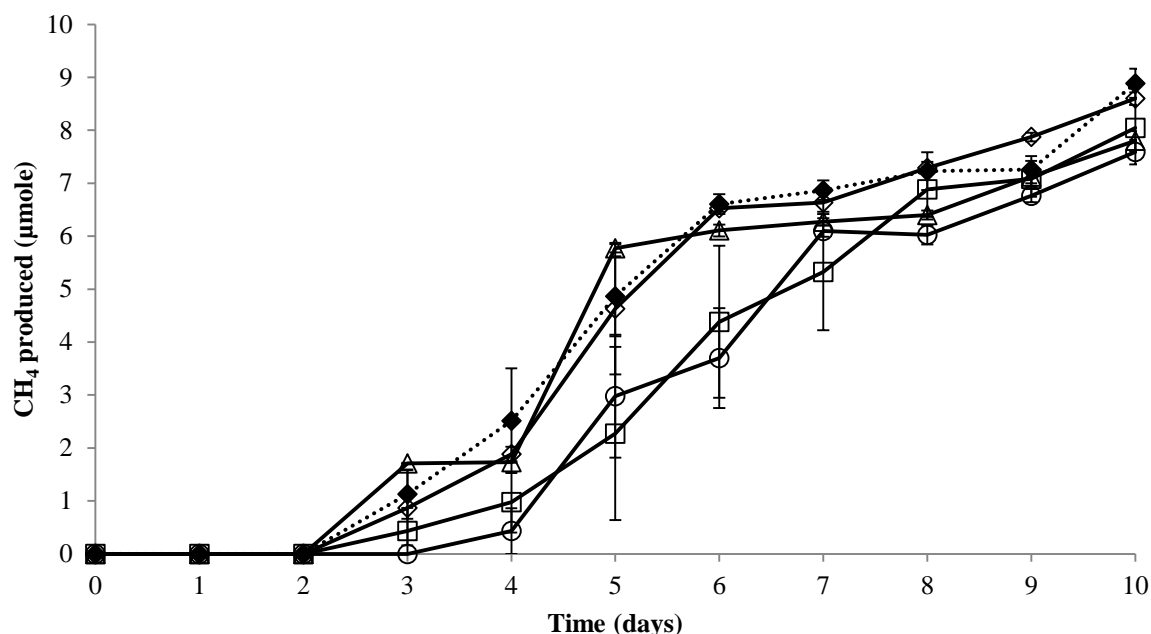
The compound associated with spectrum bin 8.71 was at low concentration, and it could not be identified from one dimensional peak alignment, TOCSY or HSQC because it was too similar to background noise. Compounds that correspond to a chemical shift of  $8.71 \pm 0.01$  include nicotinamide, nicotinurate, pyruvic acid, isoniazid, biopterin, phenanthrene, chrysene and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. Nicotinamide, a B group vitamin, was considered the most relevant to ISO4-H5 growth. The SSPGMS preparation, which contains higher concentrations of metabolites, was analysed in a prolonged  $^1\text{H}$  NMR for 72 hours to

amplify the signal strength, and compared to SSPGMS spiked with nicotinamide (300  $\mu$ M) (Figure 5.11). After comparison, the compound at chemical shift 8.71 was verified as nicotinamide.



**Figure 5.11 Identification of spectrum bin 8.71 as nicotinamide.** The spectrum of nicotinamide-spiked SSPGMS (green) was aligned with SSPGMS alone (red). Molecular structure of nicotinamide is shown at the right and the chemical shift of protons (F1) are listed in the table of assignments.

The nicotinamide concentration detected was close to the minimum detection limit of NMR, and it could not be quantified in the post-growth samples, therefore it is possibly limiting the growth of ISO4-H5. Nicotinic acid is already present in the Vitamin 10 mix added to the growth medium, and its final concentration in the medium is 0.1 mg/L. Additional nicotinamide and nicotinic acid were supplemented to ISO4-H5 culture at 100 fold that present in the standard cultivation condition, nicotinic acid was removed from the Vitamin 10 mix in the 0 mg/L nicotinic acid treatment, together with SSPGMS supplementation, and CH<sub>4</sub> production was monitored.



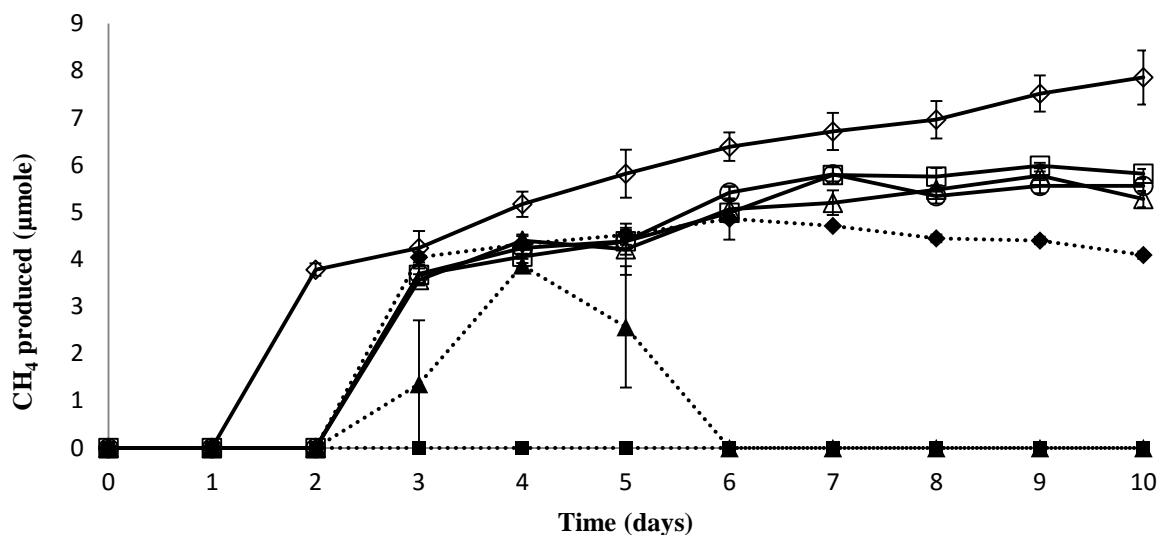
**Figure 5.12 Growth of ISO4-H5 pure culture triplicates with nicotinamide and nicotinic acid supplementation.** CH<sub>4</sub> per ml of headspace produced was recorded. All replicates contain formate (60 mM), methanol (20 mM), acetate (20 mM), 10% (v/v) SSPGMS and CoM (10 μM). (◊—) control with 0.1× vitamin mix, (Δ—) nicotinamide (0.1 mg/L) with 0.1× vitamin mix, (□—) nicotinamide (10 mg/L) with 0.1× vitamin mix, (○—) nicotinic acid (0 mg/L) with modified 0.1× vitamin mix omitting nicotinic acid, (◊••) nicotinic acid (10 mg/L) with modified 0.1× vitamin mix omitting nicotinic acid. Error bars represent ± SEMs.

The supplementation of nicotinamide or nicotinic acid in excess or removal of nicotinic acid resulted in CH<sub>4</sub> production equivalent to the control, which contains SSPGMS and vitamin mix with 0.1 mg/L nicotinic acid (Figure 5.12).

### 5.2.3 Substrate utilisation

The ISO4-H5 enrichment culture was originally grown with 20 mM methanol, and the genome analysis (Chapter 3) predicted utilization of mono-, di-, tri-methylamine and methylthiol as substrates for methanogenesis. In order to assess the utilization of alternative methyl substrates, it was first necessary to establish the minimal amount of methanol required for ISO4-H5 growth, to prevent excess methanol being carried over in the inoculum during testing of alternative methyl substrates. The reduction in methanol concentration post-growth from six replicates analysed by NMR above, ranged from 4 mM to 1.8 mM, with an average reduction of 2.6 mM. Therefore the medium used for the methanol utilisation experiment was modified to contain 4 mM methanol. CH<sub>4</sub> production from ISO4-H5 was monitored on varying concentrations of methanol, with no added methanol as negative control and 10 mM methanol as positive control (Figure 5.13). All other constituents of the medium were the same as

described in Chapter 2, except with the addition of 10 mM ethanol and 0.1 mg/L (final concentration) of nicotinamide.

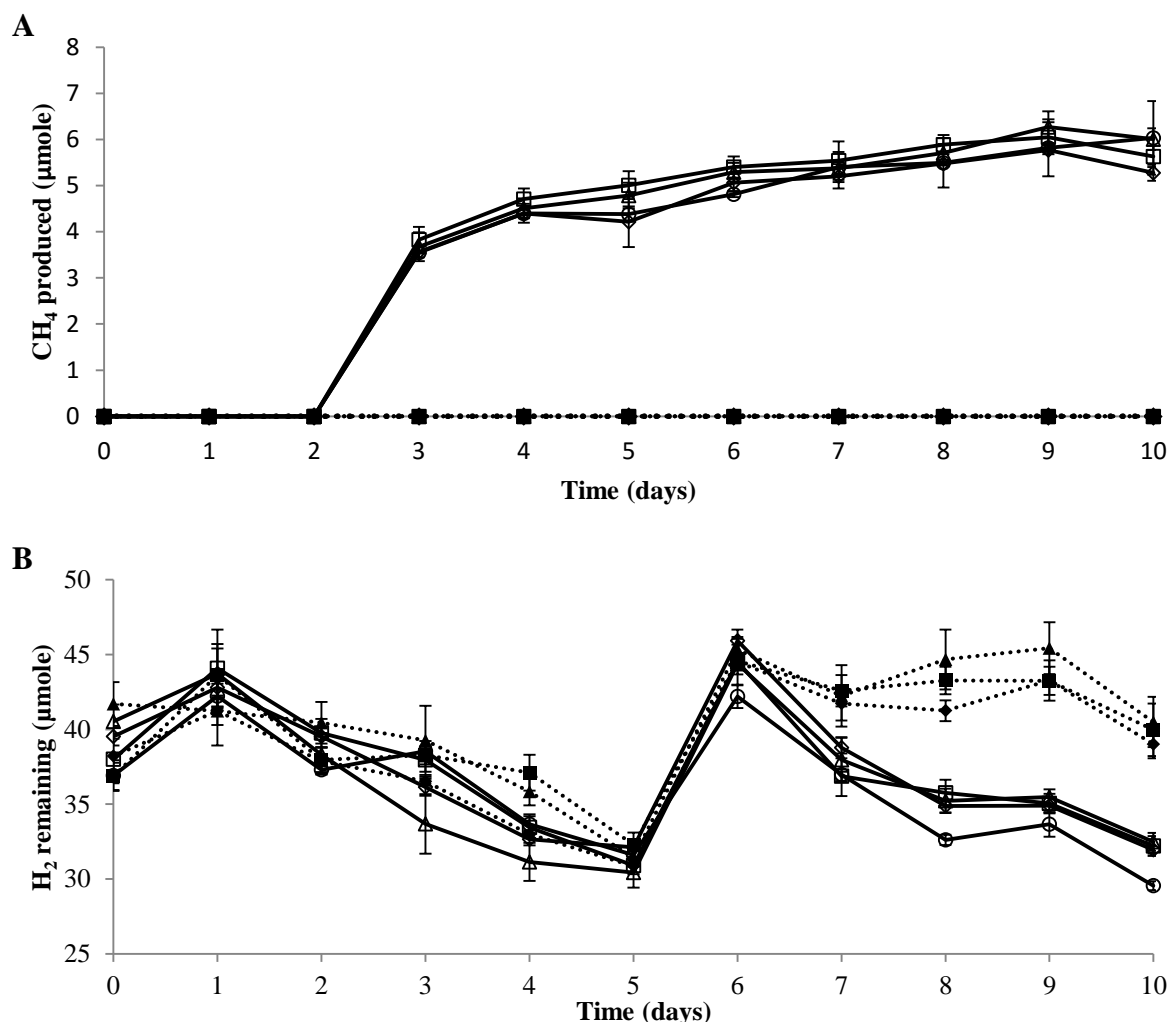


**Figure 5.13 Growth of ISO4-H5 pure culture triplicates with varying methanol concentrations.** CH<sub>4</sub> per ml of headspace produced by ISO4-H5 over 10 days. All replicates contain formate (60 mM), acetate (20 mM), CoM (10 μM) and 0.1× vitamin mix. The starter culture was grown in medium containing 4mM methanol and 10% (v/v) SSPGMS. Culture was re-pumped with H<sub>2</sub>/CO<sub>2</sub> (4:1; 200kPa) after day five. (—◇—) methanol (10 mM) with SSPGMS, (—Δ—) methanol (10 mM) without SSPGMS, (—□—) methanol (3 mM) without SSPGMS, (—○—) methanol (2 mM) without SSPGMS, (····◆····) methanol (1 mM) without SSPGMS, (····▲····) methanol (0.5 mM) without SSPGMS, (····■····) methanol (0 mM) without SSPGMS. Error bars represent ±SEMs.

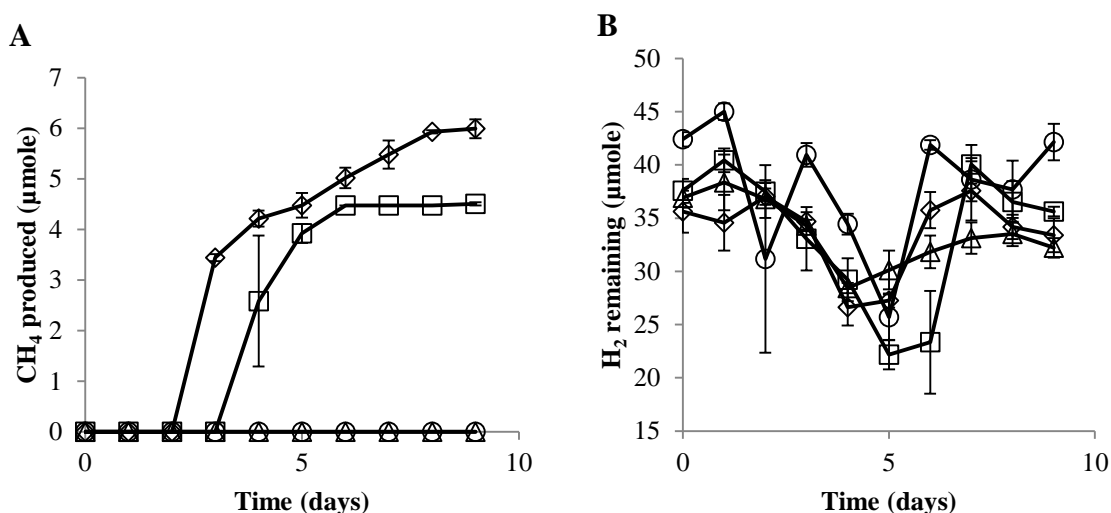
The ISO4-H5 pure culture was unable to grow without methanol (Figure 5.13A); however, it was able to grow on 0.5 mM methanol briefly before CH<sub>4</sub> production ceased at day five, the reduction of CH<sub>4</sub> on day five and six is likely due to re-pumping of the culture with H<sub>2</sub>/CO<sub>2</sub> on day five. ISO4-H5 produced less CH<sub>4</sub> than the control when grown with 1 mM methanol, with declining CH<sub>4</sub> detected after day six. There was no significant difference in CH<sub>4</sub> production from ISO4-H5 grown with 10 mM, 3 mM or 2 mM methanol. Therefore 2 mM methanol was defined as the non-limiting concentration required for ISO4-H5 growth. The ISO4-H5 culture with SSPGMS supplement produced more CH<sub>4</sub> than the cultures with 10 mM methanol (Figure 5.13A) suggesting it contains growth stimulating components in addition to the methanol, ethanol and nicotinamide previously identified by NMR analysis.

Mono-, di-, trimethylamine (10 mM), and dimethylsulfide (10 mM) were tested as possible substrates for methanogenesis, and compared to growth on methanol (10 mM) in the absence of SSPGMS (Figure 5.14). Ethanol (10 mM) and butanol (10 mM) were also tested due to

ISO4-H5 encoding a homologue of a NADP-dependent alcohol dehydrogenase (ISO4H5\_0483). The CH<sub>4</sub> production and hydrogen consumption of ISO4-H5 grown on mono-, di-, tri-methylamine were equivalent to growth on methanol (Figure 5.14A,B). ISO4-H5 was unable to produce CH<sub>4</sub> using dimethylsulfide, ethanol or butanol as substrate.

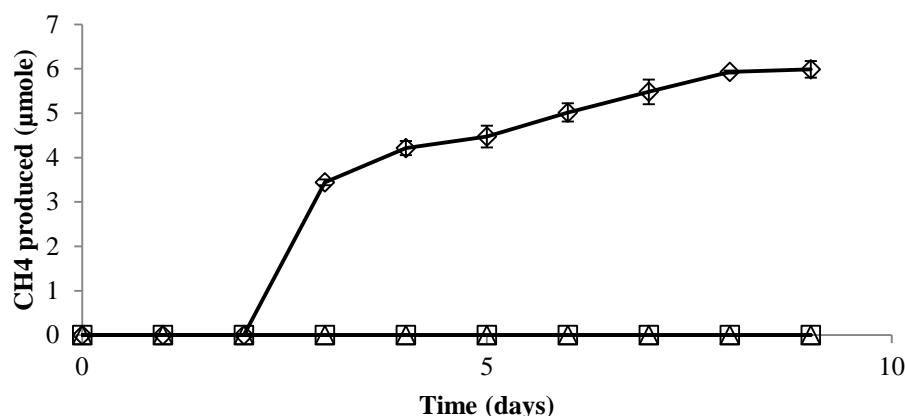


**Figure 5.14 Growth of ISO4-H5 pure culture triplicates with methyl substrates and H<sub>2</sub>/CO<sub>2</sub>.** All replicates contain formate (60 mM), acetate (20 mM), CoM (10 μM) and 0.1× vitamin mix. (—◇—) methanol (10 mM), (—Δ—) trimethylamine (10 mM), (—□—) dimethylamine (10 mM), (—○—) methylamine (10 mM), (···◆···) dimethylsulfide (10 mM), (···▲···) ethanol (10 mM), (···■···) butanol (10 mM). **A.** CH<sub>4</sub> produced per ml of headspace was measured over 10 days and expressed as μmoles of CH<sub>4</sub> per ml of headspace. **B.** Hydrogen per ml of headspace used by ISO4-H5 during growth on methyl-substrates was measured and expressed as μmoles of H<sub>2</sub> remaining per ml of headspace. Cultures were re-pumped after day 5 with H<sub>2</sub>/CO<sub>2</sub> (4:1; 200kPa). Error bars represent ±SEMs.



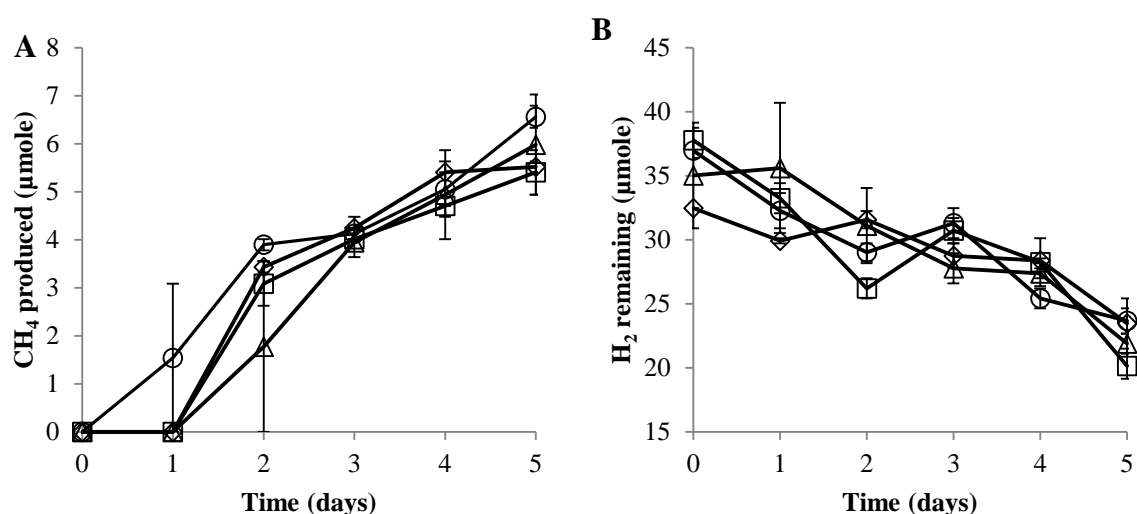
**Figure 5.15 Growth of ISO4-H5 pure culture triplicates with methyl-sources and H<sub>2</sub>/CO<sub>2</sub>.** All replicates contain formate (10 mM), acetate (20 mM), CoM (10 μM) and 0.1× vitamin mix. (—◊—) methanol (10 mM), (—Δ—) M3SP (10 mM), (—◻—) M3MSP (10 mM), (—○—) methanol (0 mM). M3SP: methylmercaptopropionate. M3MSP: methyl-3-methylthiopropionate. **A.** Amount of CH<sub>4</sub> produced per ml of headspace by ISO4-H5 over nine days. **B.** H<sub>2</sub> per ml of headspace remaining from ISO4-H5 growth on methanol. Culture was re-pumped after day five with H<sub>2</sub>/CO<sub>2</sub> (4:1; 200kPa). Error bars represents ±SEMs.

As a homologue of methylthiol:Coenzyme M methyltransferase MtsA was found in the ISO4-H5 genome, methylmercaptopropionate (M3SP) and methyl-3-methylthiopropionate (M3MSP) were also explored as possible substrates for methanogenesis. Results from these growth tests show that ISO4-H5 can use M3MSP but not M3SP. Total CH<sub>4</sub> production from M3MSP was less than from methanol and CH<sub>4</sub> formation from M3MSP was observed a day later (Figure 5.15A). In the previous growth test it was shown that ISO4-H5 does not produce CH<sub>4</sub> from ethanol. As ethanol has been reported to be able to supply reducing potential to methanogenesis in *Mbb. ruminantium* M1<sup>T</sup>, ethanol was tested in ISO4-H5 as a source of reducing potential for methanogenesis. Results from the growth test showed that ISO4-H5 cannot use ethanol to replace H<sub>2</sub> as a source of reducing potential to produce CH<sub>4</sub> (Figure 5.16).



**Figure 5.16 Growth of ISO4-H5 pure culture with putative reducing potential supplying compound in absence of H<sub>2</sub>.** All replicates contain formate (10 mM), acetate (20 mM), CoM (10 μM) and 0.1× vitamin mix. (—◇—) methanol (10 mM), (—Δ—) methanol (10 mM) and ethanol (10 mM) without H<sub>2</sub>, (—□—) methanol (10 mM) without H<sub>2</sub>. Amount of CH<sub>4</sub> produced per ml of headspace by ISO4-H5 over nine days. Culture was re-pumped after day five with H<sub>2</sub>/CO<sub>2</sub> (4:1; 200kPa). Error bars represents ±SEMs.

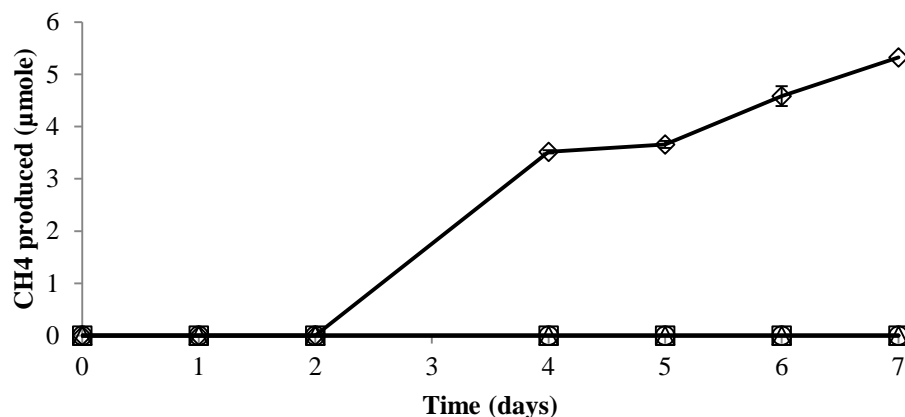
Whether ethanol supplies reducing potential in the presence of H<sub>2</sub> was also explored in growth tests. The results show that the addition of 10 mM ethanol had no effect on CH<sub>4</sub> production or hydrogen consumption (Figure 5.17A, B). Some methanogens are capable of utilising formate by oxidising it to H<sub>2</sub> and CO<sub>2</sub> via the action of formate dehydrogenase. Although no homologue of formate dehydrogenase was identified in the ISO4-H5 genome, the effect of formate on CH<sub>4</sub> production was explored. The results show that ISO4-H5 did not utilise formate to produce CH<sub>4</sub> in presence of H<sub>2</sub> (Figure 5.15), ISO4-H5 does not utilise formate in absence of H<sub>2</sub> to produce CH<sub>4</sub> (Figure 5.16), and ISO4-H5 grows equally well in high or low formate (Figure 5.17A,B).



**Figure 5.17 Effect of ethanol and formate addition on CH<sub>4</sub> formation in ISO4-H5 pure culture triplicates.** All replicates contain methanol (4 mM), acetate (20 mM), CoM (10 μM) and 0.1× vitamin mix at H<sub>2</sub>/CO<sub>2</sub> (4:1;

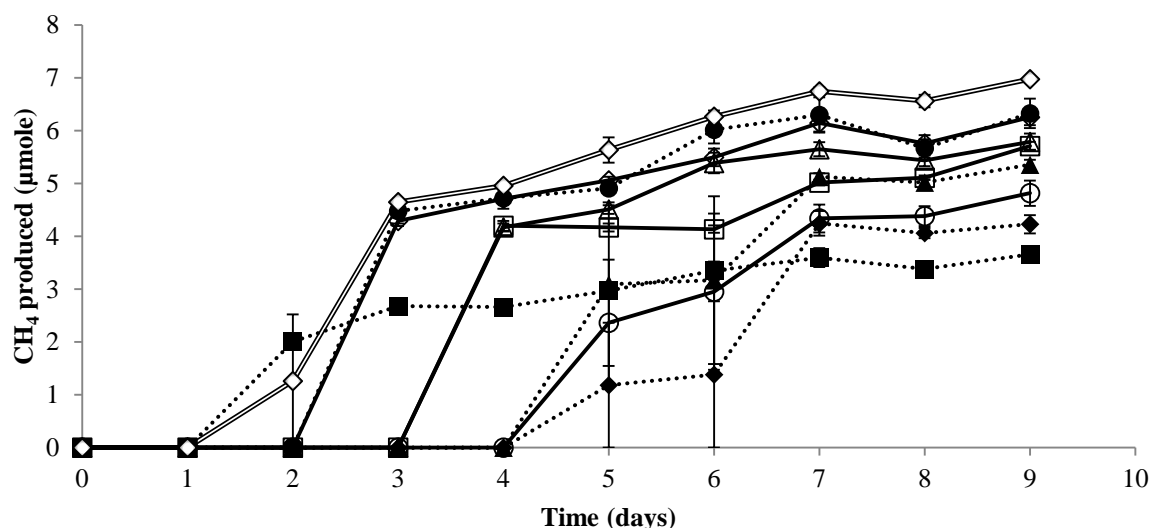


200kPa) without SSPGMS. Cultures supplemented with or without additional formate and ethanol. (—◇—) ethanol (10 mM), formate (10 mM), (—Δ—) no additional ethanol, formate, (—□—) formate (10 mM), no additional ethanol, (—○—) ethanol (10 mM), no additional formate. There is a residual of 0.01 mM ethanol and 0.06 mM formate carried over from previous inoculations. **A.** Amount of CH<sub>4</sub> produced per ml of headspace by ISO4-H5 over five days. **B.** Hydrogen per ml of headspace remaining from ISO4-H5 growth on methanol over five days. Error bars represents ±SEMs.



**Figure 5.18 Testing growth of ISO4-H5 pure culture triplicates on betaine or choline and H<sub>2</sub>/CO<sub>2</sub> (4:1; 200kPa).** All replicates contain formate (60 mM), methanol ( $\leq 0.4$  mM) carried over from previous inoculation, acetate (20 mM), CoM (10  $\mu$ M) and 0.1 $\times$  vitamin mix. (—◇—) Methanol (10 mM), (—Δ—) Betaine (10 mM), (—□—) Choline (10 mM), (—○—) Methanol (0 mM). Amount of CH<sub>4</sub> produced per ml of headspace by ISO4-H5 over seven days. Error bars represents ±SEMs.

The precursors of methylamines in the rumen are likely to be choline and/or betaine, therefore these compounds were explored as possible substrates for methanogenesis by ISO4-H5. Results from these growth tests showed that ISO4-H5 was unable to produce CH<sub>4</sub> using betaine or choline (Figure 5.18).



**Figure 5.19 ISO4-H5 pure culture supplementation.** Pure culture of ISO4-H5 supplemented with amino acids and other metabolites in triplicates. All replicates contain formate (60 mM), acetate (20 mM), methanol (20 mM), CoM (10  $\mu$ M), 10% (v/v) SSPGMS and 0.1 $\times$  vitamin mix. (—◇—) Control, (—Δ—) NaCl (1% w/v), (—□—) succinate

(10 mM), (—○—) citrate (10 mM), (···◆···) malate (10 mM), (···▲···) fumarate (10 mM), (···■···) aspartate (10 mM), (···●···) glycine (10 mM), (=◆=) serine (10 mM). Amount of CH<sub>4</sub> produced per ml of headspace by ISO4-H5 over nine days. Error bars represents  $\pm$ SEMs.

As the growth of ISO4-H5 *in vitro* was slow and sub-optimal, various supplements and metabolic intermediates were explored to identify growth limiting factors (Figure 5.19). The betaine and choline were tested separately from other supplements due to requiring a valid negative control, which the carried-over methanol concentration is lesser than what ISO4-H5 required to survive, while other supplements were tested in presence of methanol as a methyl source. *Methanomassiliicoccus luminyensis* was reported as having optimal growth at higher NaCl concentrations, therefore 1% NaCl (w/v) was tested. The ISO4-H5 genome is predicted to encode an incomplete citric acid cycle, therefore succinate, citrate, malate, fumarate and aspartate (all at 10 mM) were tested as growth supplements, along with glycine and serine, as ISO4-H5 was predicted to lack a serine hydroxymethyltransferase gene required for glycine biosynthesis from serine. Results from these growth tests showed that ISO4-H5 does not produce more CH<sub>4</sub> with NaCl, succinate, citrate, malate, fumarate, serine or glycine supplementation (Figure 5.19). While aspartate supplementation allowed earlier CH<sub>4</sub> formation in ISO4-H5, overall CH<sub>4</sub> production was lower than the unsupplemented control.

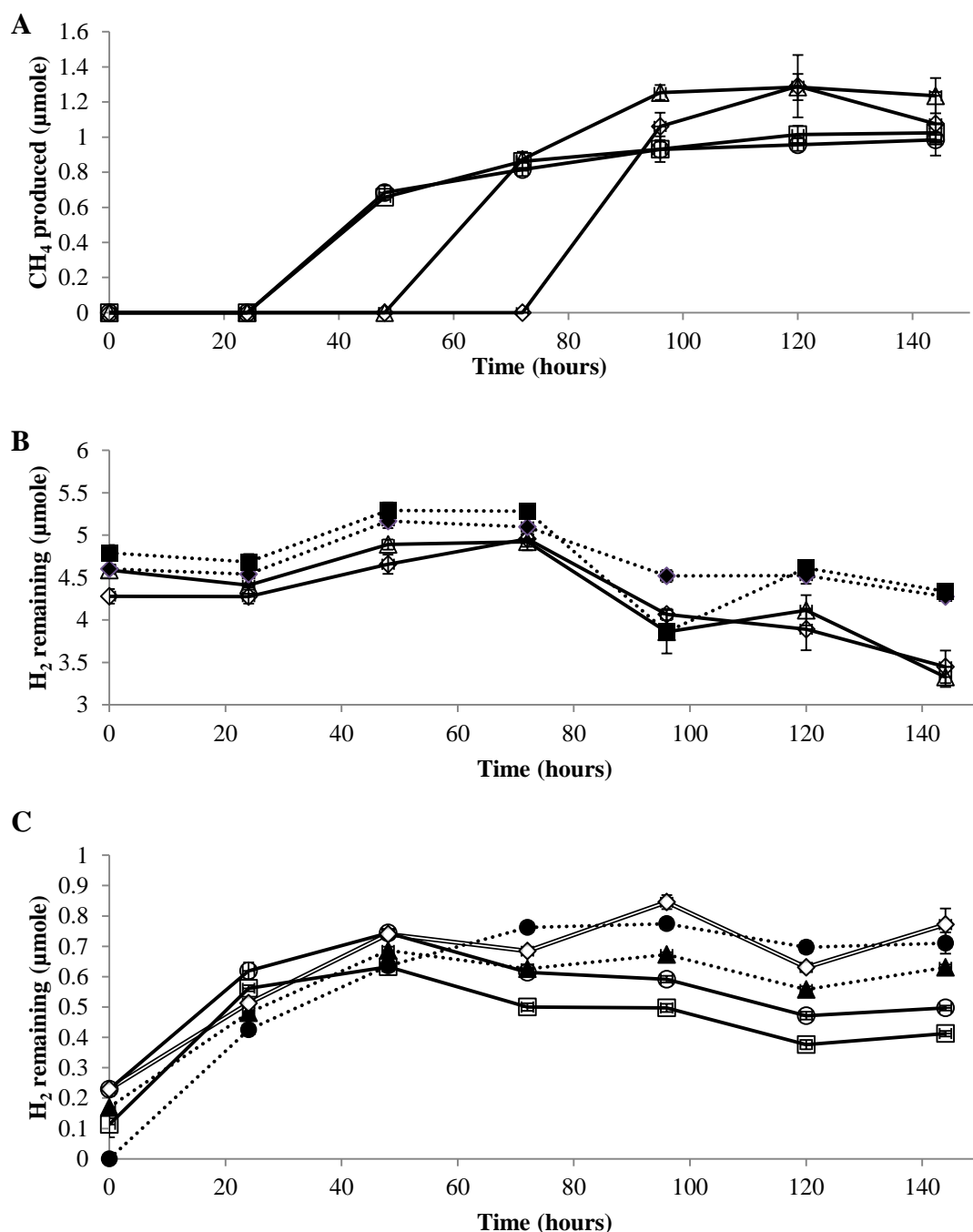
#### 5.2.4 Analysis of ISO4-H5 gene expression during growth on different methanogenic substrates and under high or low H<sub>2</sub>

Methanogen growth is dependent on the concentration of hydrogen and the type of methyl substrate, therefore it was of interest to examine ISO4-H5 gene expression under conditions varying in hydrogen level and methyl substrate type. An experiment was conducted using the ISO4-H5 enrichment culture grown under high and low hydrogen conditions. The high hydrogen condition was provided using a headspace of H<sub>2</sub>/CO<sub>2</sub> (4:1; 200kPa), while the low hydrogen environment was simulated by co-culture with the cellulose-degrading, hydrogen-producing bacterium, *Ruminococcus flavefaciens* FD1 grown on cellulose with a N<sub>2</sub> (200kPa) atmosphere headspace (Figure 5.20C). The experimental set up of the co-cultures is summarised in Table 5.3.

**Table 5.3 Experimental set up of high and low hydrogen co-cultures of ISO4-H5 enrichment grown with methanol or methylamine.**

Treatment	Co-culture description	Methanol	Methylamine	H <sub>2</sub>	FD1*	Sdex*	ISO4-H5 enrichment
1	ISO4-H5 high H <sub>2</sub> + methanol	20 mM	-	High	-	-	+
2	ISO4-H5 + FD1 high H <sub>2</sub> + methanol	20 mM	-	High	+	-	+
3	ISO4-H5 + FD1 low H <sub>2</sub> + methanol	20 mM	-	Low	+	-	+
4	ISO4-H5 + FD1 low H <sub>2</sub> + methylamine	-	20 mM	Low	+	-	+
5	Sdex + FD1 high H <sub>2</sub>	20 mM	-	High	+	+	-
6	Sdex + FD1 low H <sub>2</sub>	20 mM	-	Low	+	+	-
7	FD1 high H <sub>2</sub>	20 mM	-	High	+	-	-
8	FD1 low H <sub>2</sub>	20 mM	-	Low	+	-	-
9	Sdex + FD1 low H <sub>2</sub> + methylamine	-	20 mM	Low	+	+	-

\*FD1: *Ruminococcus flavefaciens* FD1. Sdex: *Succinivibrio dextrinosolvens* H5. Each culture was supplemented with 20 mM acetate, 1 mM CoM, 0.1× Vitamin 10 mix and 0.2% (w/v) cellulose.



**Figure 5.20 Growth of ISO4-H5 enrichment culture under high or low hydrogen with either methanol or methylamine.** (—◇—) ISO4-H5 high H<sub>2</sub> + methanol, (—Δ—) ISO4-H5 + FD1 high H<sub>2</sub> + methanol, (—□—) ISO4-H5 + FD1 low H<sub>2</sub> + methanol, (—○—) ISO4-H5 + FD1 low H<sub>2</sub> + methylamine, (···◆···) Sdex + FD1 high H<sub>2</sub>, (···▲···) Sdex + FD1 low H<sub>2</sub>, (···■···) FD1 high H<sub>2</sub>, (···●···) FD1 low H<sub>2</sub>, (—◇=) Sdex + FD1 low H<sub>2</sub> + methylamine. **A.** Amount of CH<sub>4</sub> produced per ml of headspace in ISO4-H5 enrichment culture under high or low hydrogen conditions. **B.** Amount of H<sub>2</sub> per ml of headspace consumed by ISO4-H5. **C.** Amount of H<sub>2</sub> per ml of headspace produced by FD1 and consumed by ISO4-H5. Error bars represent  $\pm$ SEMs.

CH<sub>4</sub> production was observed earliest when ISO4-H5 was grown on methanol under the FD1 + low hydrogen condition but a higher final CH<sub>4</sub> concentration was observed in the stationary phase ISO4-H5 co-culture with FD1 under high hydrogen conditions (Figure 5.20A). Approximately one  $\mu$ mole of hydrogen per ml of headspace was consumed by ISO4-H5 under

high hydrogen conditions (Figure 5.19B). The hydrogen produced from FD1 appears to be more than sufficient for ISO4-H5 growth, as accumulation of hydrogen under low hydrogen conditions was observed (Figure 5.20C). The FD1 monoculture produced more hydrogen than the co-culture of FD1 with *S. dextrinosolvens* H5 (Figure 5.20C). ISO4-H5 grown on methanol and methylamine produced equivalent amounts of CH<sub>4</sub> under the low hydrogen condition (Figure 5.20A).

RNA extraction was carried out on culture samples taken at 120 h post-inoculation, as described in Chapter 2. Sufficient quantity of RNA for sequencing could only be collected from the cultures containing ISO4-H5. The quality of RNA in each replicate is displayed in Table 5.4. Three of the four replicates were transcriptome sequenced.

**Table 5.4 Quality of RNA extracted from ISO4-H5 co-cultures**

RNA sample	RNA integrity number
1-A (120 h)	N/A
1-C (120 h)	9.3
1-E (120 h)	9.2
1-F (120 h)	9.1
2-A (120 h)	9.3
2-C (120 h)	9.8
2-E (120 h)	8.9
2-H (120 h)	9.1
3-A (120 h)	8.9
3-B (120 h)	8.8
3-D (120 h)	8.8
3-H (120 h)	9
4-A (120 h)	9
4-B (120 h)	9
4-E (120 h)	9.2
4-F (120 h)	9.1

To calculate gene expression levels, the transcripts were trimmed and assembled to the ISO4-H5 reference genome at 98% similarity via the Rockhopper and Edge-pro pipelines (Table 5.5). In a preliminary test of these two assemblers, the number of transcripts aligned to ISO4-H5 by Edge-pro was higher compared to Rockhopper, however, the Edge-pro alignment could not differentiate transcripts from the three co-cultured organisms, therefore Rockhopper was chosen for this task.

**Table 5.5 Reads aligned to ISO4-H5, *S. dextrinosolvens* H5 and *R. flavefaciens* FD1 genomes using Rockhopper.**

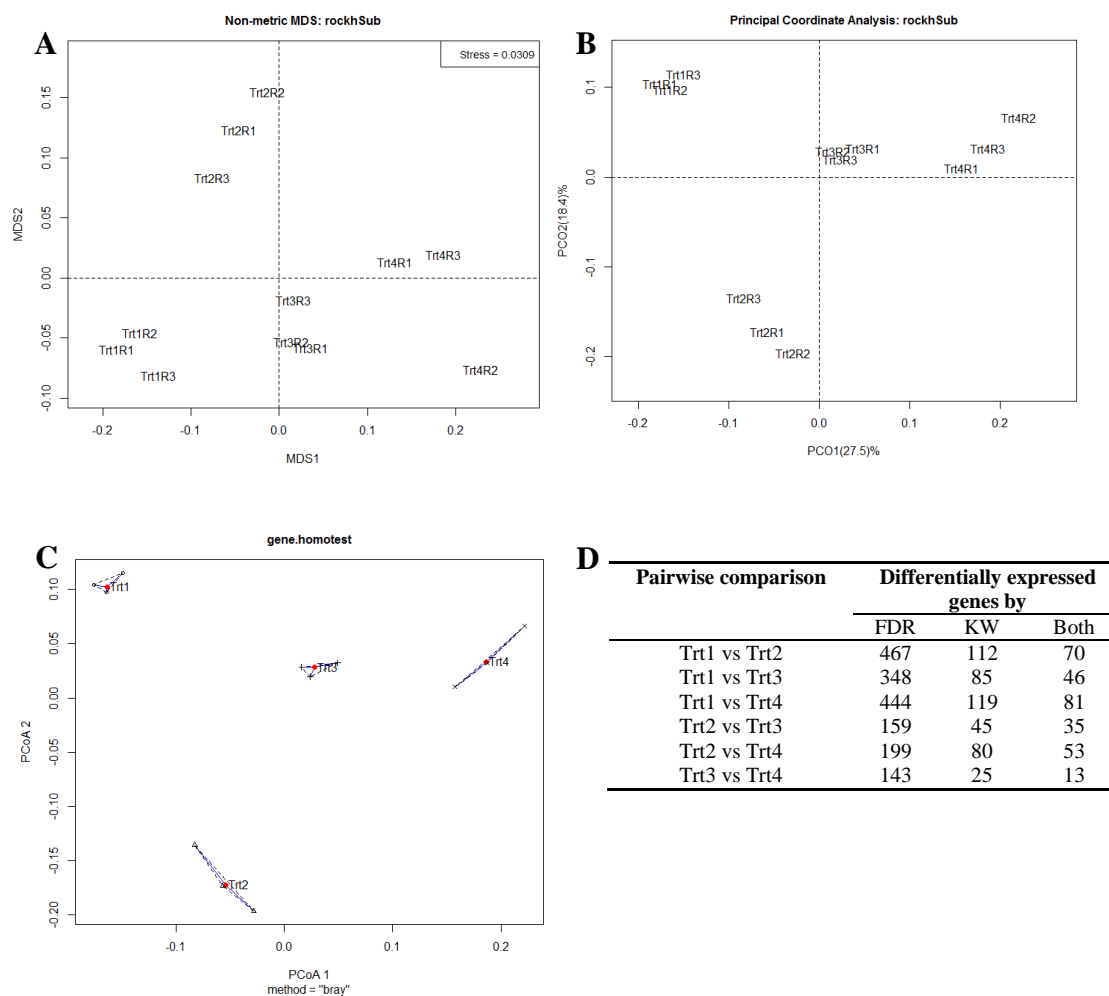
Treatment description	Replicate	Total reads	Reads aligned to					
			ISO4-H5	% total	Sdex	% total	FD1	% total
Trt1: ISO4-H5 high H <sub>2</sub> + methanol	1	9574814	7188267	75%	674275	7%		
	2	11171088	7396072	66%	1726545	15%		
	3	12762460	8004660	63%	2191541	17%		
Trt2: ISO4-H5 + FD1 high H <sub>2</sub> + methanol	1	8907622	5000664	56%	1266430	14%	1119464	13%
	2	8082221	5068906	63%	847775	10%	855331	11%
	3	9896974	6586875	67%	672412	7%	788202	8%
Trt3: ISO4-H5 + FD1 low H <sub>2</sub> + methanol	1	8704366	6066984	70%	480305	6%	376550	4%
	2	10191472	6643413	65%	613478	6%	757425	7%
	3	8012444	5698164	71%	457562	6%	448399	6%
Trt4: ISO4-H5 + FD1 low H <sub>2</sub> + methylamine	1	11000283	8689292	79%	643327	6%	51385	0.5%
	2	9758529	7610170	78%	646486	7%	99859	1%
	3	11799619	9246903	78%	860353	7%	62816	1%

### 5.2.5 Statistical analyses

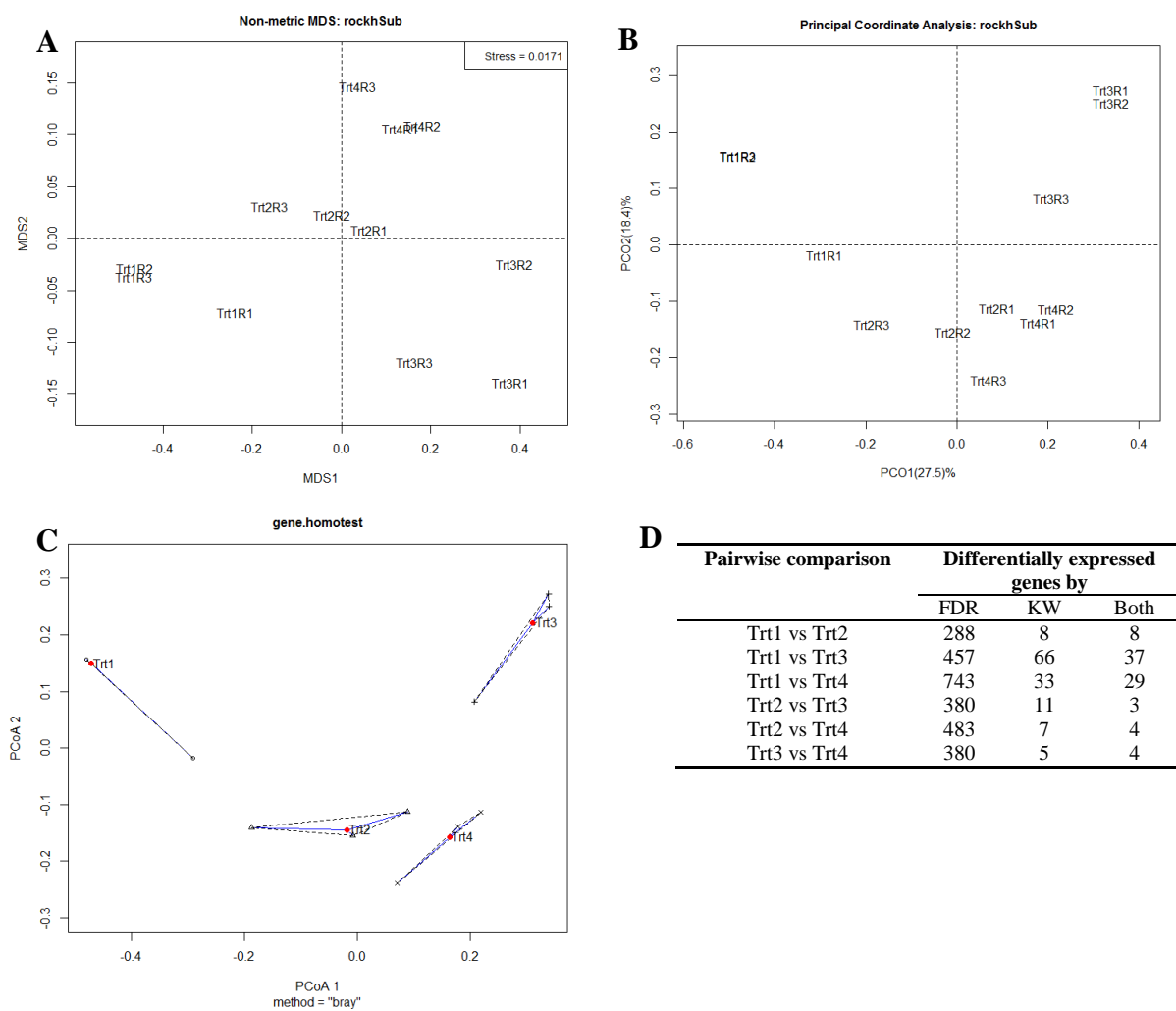
The read counts of mapped transcripts were normalized and differential expression analyses were performed by Rockhopper, using a false discovery rate (FDR) cutoff of <0.05, a non-parametric Kruskal-Wallis rank sum test (KW) *p*-value cutoff of <0.05, as well as fold change >2, to define differentially expressed genes between any two conditions. Gene expression profiles of the three genomes are displayed in Tables A.5.3, A.5.4 and A.5.5.

To investigate the difference between conditions as well as replicates, multivariate analyses were performed on genes differentially expressed between two conditions (Figures 5.21, 5.22, 5.23). The MDS plot of differentially expressed genes in ISO4-H5 showed clustering of replicates within Trt1, Trt2 and Trt3 and more dissimilarity was observed between replicates of Trt4. The *S. dextrinosolvens* H5 MDS plot showed similarity between replicates of Trt2 and Trt4, and replicate one of Trt1 appeared less similar to the two other replicates. The MDS plot of differentially expressed genes in *R. flavefaciens* FD1 displayed strong dissimilarity among replicates of Trt4.

The PCoA plots of differentially expressed genes in ISO4-H5 and *R. flavefaciens* FD1 showed clustering of replicates within treatments, whereas with *S. dextrinosolvens* H5 replicates of Trt2 and Trt4 do not display obvious clustering and is similar even between treatments, while replicate 1 of Trt1 and replicate 3 of Trt3 displayed dissimilarity to other replicates within the treatment. The homogeneity plot of differentially expressed genes in ISO4-H5 and *R. flavefaciens* FD1 showed low transcriptional diversity within treatments, whereas larger dispersion within treatments was observed in *S. dextrinosolvens* H5.

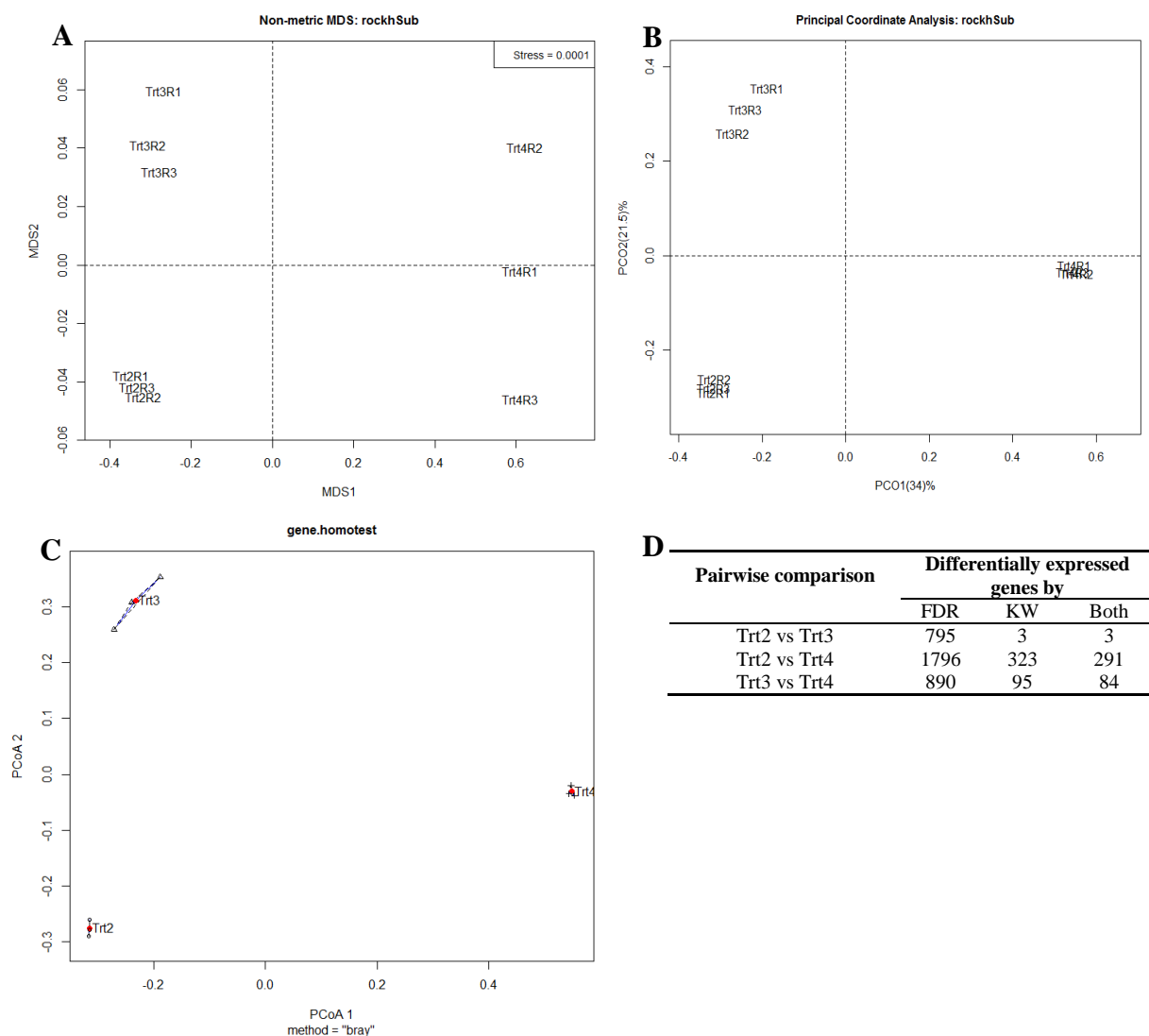


**Figure 5.21** Multivariate analysis based on relative abundance of transcripts in the ISO4-H5 transcriptome via Bray Curtis distance matrix. **A.** Nonmetric multidimensional scaling plot **B.** Principal coordinate analysis plot. **C.** Group dispersions plot by permutation-based test of multivariate homogeneity. **D.** Table of numbers of differentially expressed genes determined by FDR and KW. Trt1-4 are as described in Table 5.3.



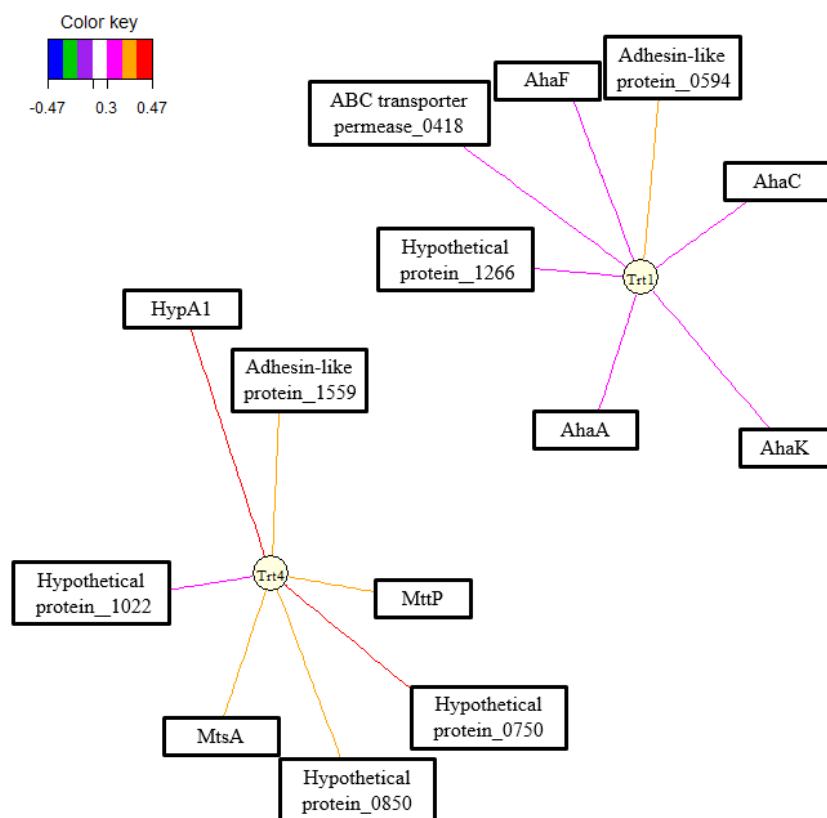
**Figure 5.22 Multivariate analysis based on relative abundance of transcripts in the *S. dextrinosolvens* H5 transcriptome via Bray Curtis distance matrix. A.** Nonmetric multidimensional scaling plot. **B.** Principal coordinate analysis plot. **C.** Group dispersions plot by permutation-based test of multivariate homogeneity. **D.** Table of differentially expressed genes. Trt1-4 are as described in Table 5.3.





**Figure 5.23** Multivariate analysis based on relative abundance of transcripts in the *R. flavefaciens* FD1 transcriptome via Bray Curtis distance matrix. **A.** Nonmetric multidimensional scaling plot. **B.** Principal coordinate analysis plot. **C.** Group dispersions plot by permutation-based test of multivariate homogeneity. **D.** Table of differentially expressed genes. Trt1-4 are as described in Table 5.3.

Correspondence analysis was also used to explore the association between treatments and differentially expressed genes in ISO4-H5. At a CA threshold of 0.3, only genes positively associated with Trt1 and Trt4 were observed (Figure 5.24).



**Figure 5.24 Network diagram showing the associations between genes and treatments based on Correspondence Analysis.** The threshold used was 0.3. Solid lines represents positive association, dotted lines represents negative association. Line colours signify magnitude of association as illustrated by the key at top left. A<sub>1</sub>A<sub>0</sub> ATP synthase subunits (Aha). Trimethylamine permease (MttP). Methylthiol:corrinoid methyltransferase (MtsA). Hydrogenase nickel insertion protein (HypA1).

#### 5.2.6 Differential gene expression in ISO4-H5

There were 298 differentially expressed ISO4-H5 genes that fit the aforementioned statistical cut-offs (Figure 5.21D, Table A.5.3). In order to assess the ISO4-H5 response to H<sub>2</sub> level while using methanol, the 35 genes differentially expressed between Trt2 (high H<sub>2</sub>) and Trt3 (low H<sub>2</sub>) were investigated first (Table 5.6).

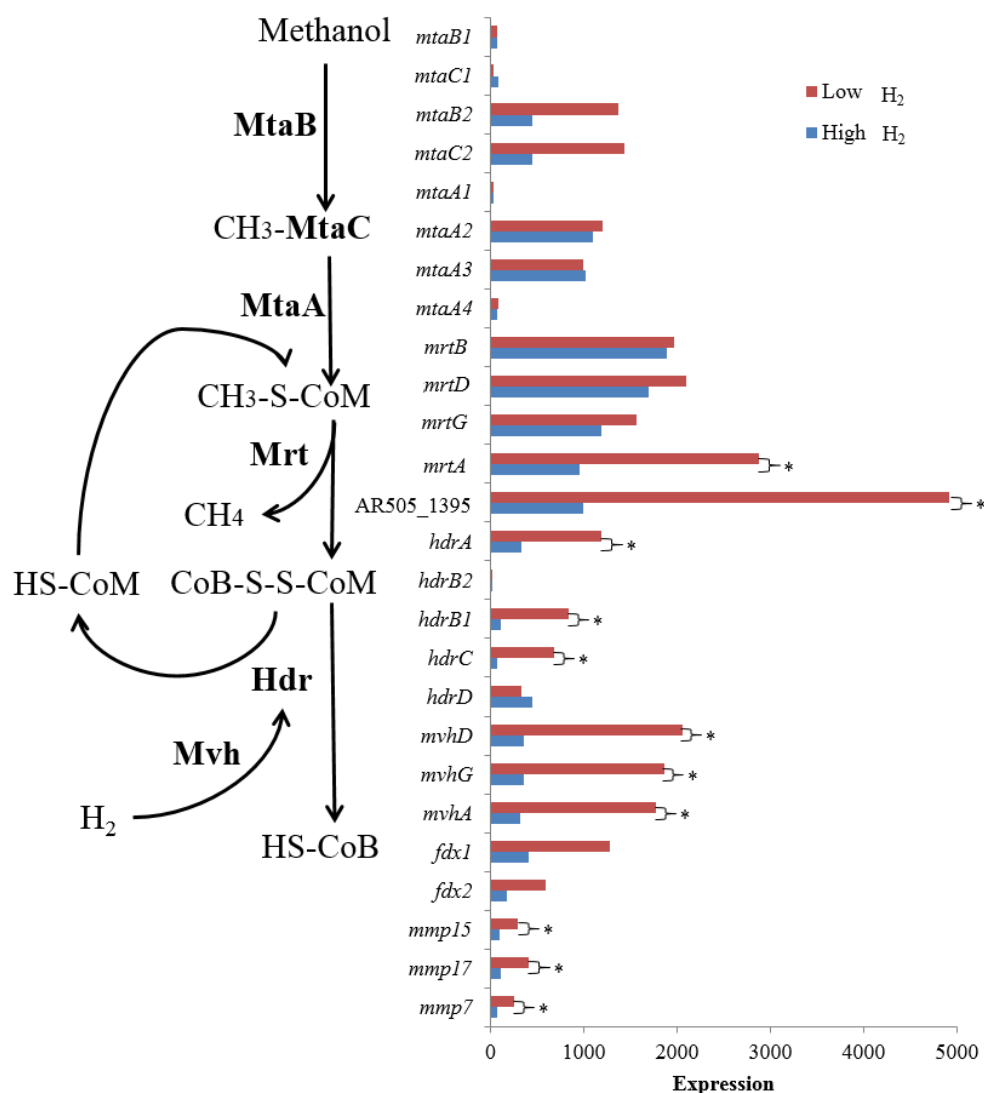
**Table 5.6 ISO4-H5 genes differentially expressed between Trt2 (high H<sub>2</sub>) and Trt3 (low H<sub>2</sub>)**

Locus_tag	Product	Expression (RPK0.1M)			
		Trt1	Trt2	Trt3	Trt4
AR505_0038	ribosomal RNA large subunit methyltransferase J RrmJ	144	245	113	140
AR505_0126	transmembrane protein	118	172	56	92
AR505_0199	transcriptional regulator AsnC family	72	45	101	45
AR505_0225	ribosomal protein L29P Rpl29p	235	81	301	177
AR505_0226	translation initiation factor aSUI1	298	65	358	163
AR505_0234	ribosomal protein L6P Rpl6p	168	51	178	94
AR505_0273	CoB-CoM heterodisulfide reductase subunit C HdrC	93	79	681	166
AR505_0274	CoB-CoM heterodisulfide reductase subunit B HdrB1	182	110	836	279
AR505_0275	TPR repeat-containing protein	13	8	38	16
AR505_0518	hypothetical protein	37	20	45	21
AR505_0536	TPR repeat-containing protein	122	195	88	125
AR505_0675	acetylglutamate kinase ArgB	29	15	48	34
AR505_0689	transmembrane protein	150	235	104	148
AR505_0909	adhesin-like protein	128	263	114	164
AR505_1118	MTA/SAH nucleosidase MtnN	35	61	26	45
AR505_1155	quinolinate synthetase A protein NadA	113	196	98	186
AR505_1187	hypothetical protein	221	1091	159	474
AR505_1188	transmembrane protein	67	125	45	102
AR505_1216	TPR repeat-containing protein	63	98	42	51
AR505_1233	pseudouridylate synthase	48	87	38	63
AR505_1351	TPR repeat-containing protein	51	92	40	59
AR505_1386	methanogenesis marker protein 17	381	118	411	209
AR505_1387	methanogenesis marker protein 15	256	95	298	136
AR505_1388	methanogenesis marker protein 5	101	48	131	53
AR505_1395	hypothetical protein	1683	999	4926	1831
AR505_1396	methyl-coenzyme M reductase alpha subunit McrA	1696	953	2883	1328
AR505_1476	methyl viologen-reducing hydrogenase alpha subunit MvhA	436	320	1771	861
AR505_1477	methyl viologen-reducing hydrogenase gamma subunit MvhG	515	364	1870	1014
AR505_1478	methyl-viologen-reducing hydrogenase delta subunit	400	361	2055	1072
AR505_1479	CoB-CoM heterodisulfide reductase subunit A HdrA	381	327	1194	748
AR505_1583	tRNA nucleotidyltransferase Cca	68	138	54	108
AR505_1648	transketolase subunit B	120	65	170	87
AR505_1661	hypothetical protein	87	170	56	88
AR505_1675	radical SAM domain-containing protein	85	114	50	89
AR505_1782	DNA primase large subunit PriB	128	187	93	102

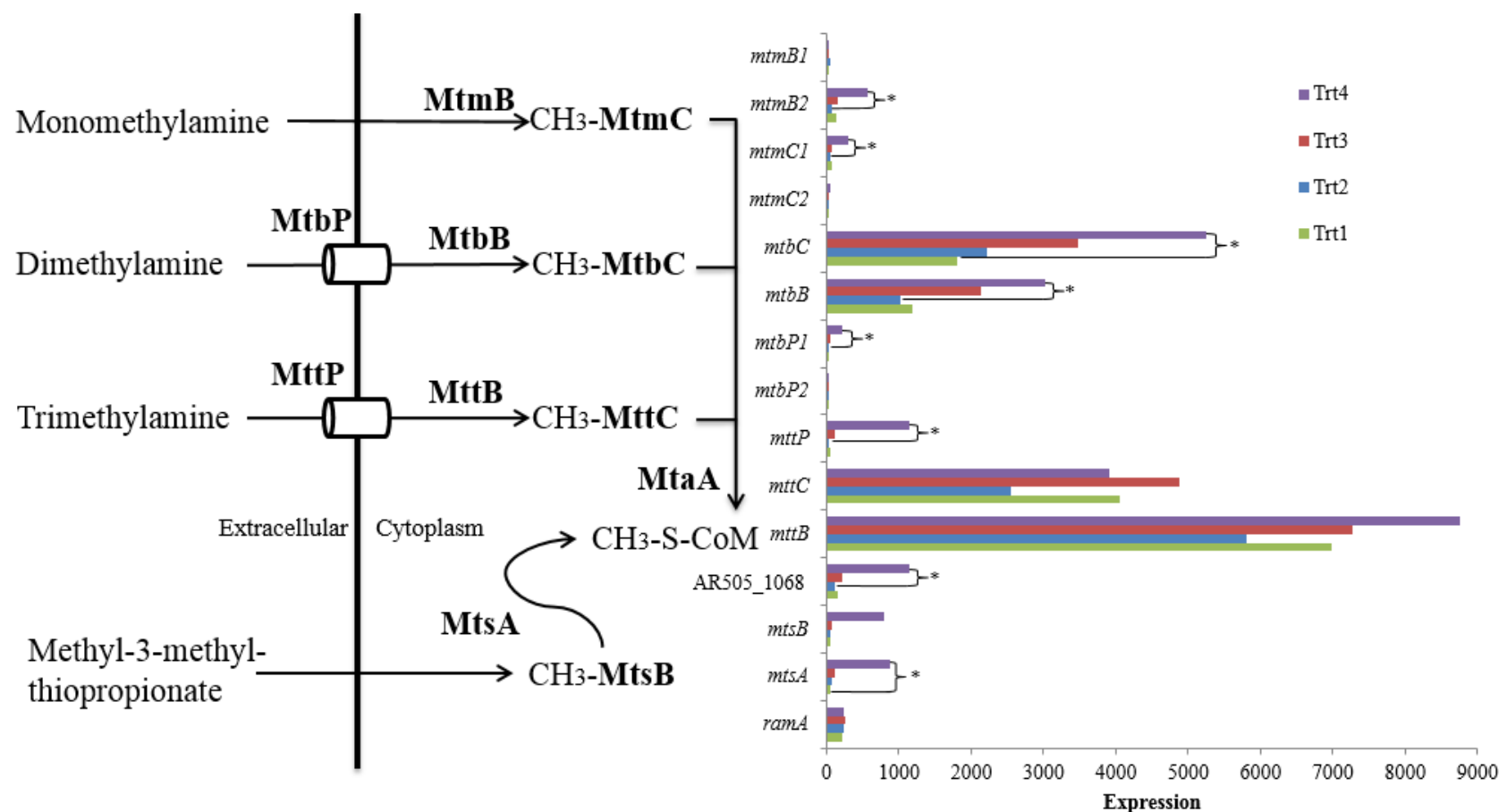
The expression of these genes and the location of their corresponding enzymes in the methylotrophic methanogenesis pathway in ISO4-H5, are summarised in Figure 5.25. There were 19 genes with higher expression under low H<sub>2</sub> compared to high H<sub>2</sub>, including *mcrA* (3× up-regulated) which is central to methanogenesis, *hdrA* (3×), *hdrB* (8×), *hdrC* (9×) involved in cofactor regeneration and *mvhAGD* (5×) involved in supplying reducing potential for methanogenesis. Four additional genes associated with methanogenesis also had higher expression under low H<sub>2</sub>: AR505\_1395 (encoding a hypothetical protein immediately downstream of *mcrA*, 5×) and 3 genes encoding methanogenesis marker proteins (MMP) *mmp5*, *mmp15* and *mmp17* (3×). In addition, one set of methanol utilizing genes (*mtaB2*, *mtaC2*) and a pair of ferredoxins genes *fdx1* and *fdx2* (AR505\_0452, AR505\_0453) have 3× higher expression under low H<sub>2</sub> which meet the FDR cutoff (*mtaB2*:  $3.2 \times 10^{-4}$ , *mtaC2*:  $5.3 \times 10^{-5}$ , *fdx1*:  $1.4 \times 10^{-6}$ , *fdx2*:  $2.0 \times 10^{-7}$ ) but not the KW *p* value cutoff (*mtaB2*: 0.22, *mtaC2*: 0.14, *fdx1*: 0.14, *fdx2*: 0.14). Among the 16 genes that showed higher expression under high

H<sub>2</sub>, hypothetical protein AR505\_1187 had the highest fold change of 7×. Other genes differentially expressed were involved in replication and protein synthesis, a *nadA* gene involved in NAD biosynthesis from aspartate, genes presumed to be involved in regulation and yet other genes with unknown functions (Table 5.6).

In order to assess the ISO4-H5 response to different methanogenic substrates, growth on methylamine (Trt4) was compared with methanol (Trt3). Only 13 genes were differentially expressed in this comparison, including genes encoding three Fpo-like complex proteins (*fpoN*, *fpoM* and *fpoL*). No genes directly associated with methanogenic substrate utilisation were differentially expressed under these conditions. Therefore, the genes encoding methylamine and methylthiol utilisation were investigated across all four treatments (Figure 5.25). Higher expression of genes encoding methylamine use, such as *mtmB2* (8×), *mtmC1* (6×), dimethylamine utilising *mtbC* (3×), *mtbB* (3×), *mtbP1* (13×) and trimethylamine transporting *mttP* (42×) as well *mtsA* encoding methylthiol use, were observed in Trt4 (methylamine under low H<sub>2</sub>), in comparison to either Trt1 or Trt2 (Figure 5.26, Table A.5.3). Genes other than those encoding the methanogenesis pathway may also be involved in adaptation to growth on different methanogenic substrates, therefore genes with two-fold difference in expression in Trt4 compared to all three other treatments were investigated and are summarised in Table 5.07. Eight out of 11 subunits of the Fpo-like complex were found to have low expression in Trt4. However, the expression of Fpo-like complex genes is likely influenced by multiple factors, as illustrated in Figure 5.28A.



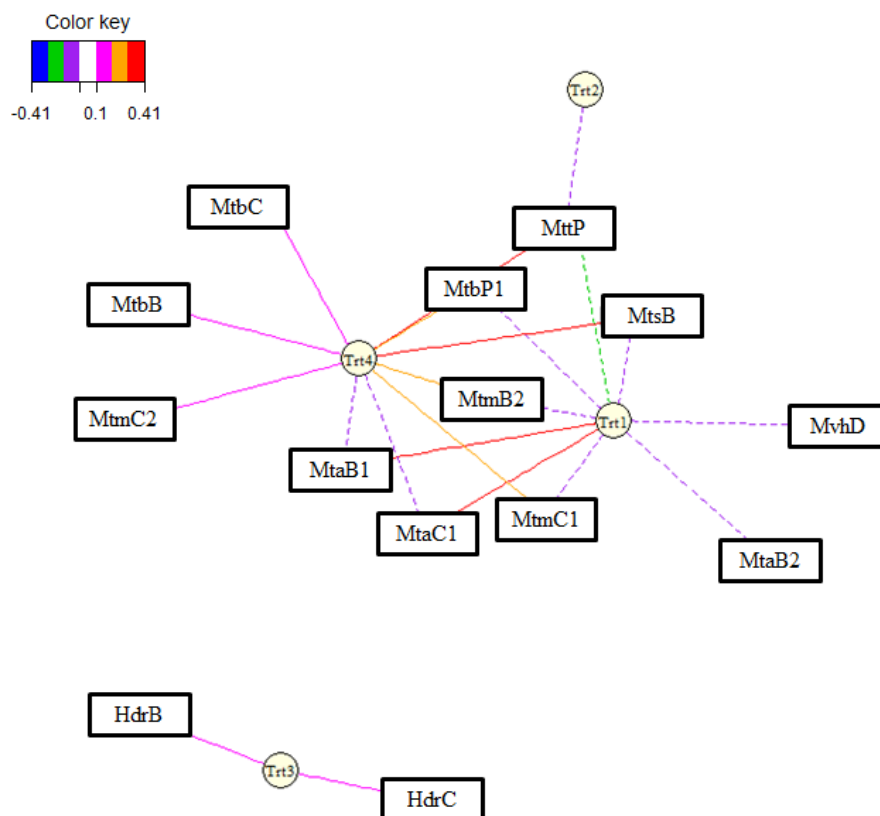
**Figure 5.25 Expression of genes involved in methanol utilizing methanogenesis under high and low H<sub>2</sub> conditions.** \*: Differentially expressed genes with FDR <0.05, KW  $p$  <0.05, fold change >2 between Trt2 (high H<sub>2</sub>) and Trt3 low H<sub>2</sub>). Methanol:corrinoid methyltransferase (MtaB), methanol corrinoid protein (MtaC), Methyl:coenzyme M methyltransferase (MtaA), methyl-coenzyme M reductase complex (Mcr), hypothetical protein immediate downstream of *mcrA* gene (AR505\_1395), heterodisulfide reductase (Hdr), methyl viologen hydrogenase (Mvh), ferredoxin (Fdx), methanogenesis marker protein (MMP). Gene products are displayed in bold within the pathway.



**Figure 5.26 Expression of ISO4-H5 genes involved in methylamine and methylthiol use.** \*: Differentially expressed genes with FDR <0.05, KW p <0.05, fold change >2. Trt1: ISO4-H5 enrichment culture on methanol with high H<sub>2</sub>. Trt2: ISO4-H5 enrichment culture +FD1 on methanol with high H<sub>2</sub>. Trt3: ISO4-H5 enrichment culture +FD1 on methanol with low H<sub>2</sub>. Trt4: ISO4-H5 enrichment culture +FD1 on monomethylamine with low H<sub>2</sub>. Monomethylamine:corrinoid methyltransferase (MtmB), monomethylamine corrinoid protein (MtmC), dimethylamine:corrinoid methyltransferase (MtbB), dimethylamine corrinoid protein (MtbC), dimethylamine permease (MtbP). Trimethylamine:corrinoid methyltransferase (MttB), trimethylamine corrinoid protein (MttC), trimethylamine permease (MttP), 4Fe-4S ferredoxin iron-sulfur binding domain-containing protein (AR505\_1068), methylthiol corrinoid protein (MtsB), bifunctional methylthiol:corrinoid methyltransferase (MtsA), methylamine methyltransferase corrinoid activation protein (RamA). Gene products are displayed in bold within the pathway. Treatment groups are defined in Table 5.3.

**Table 5.7 Genes with 2-fold expression difference in Trt4 compared with other treatments**

Locus_tag	Product	Expression (RPK0.1M)			
		Trt1	Trt2	Trt3	Trt4
Genes with 2-fold higher expression					
AR505_0443	universal stress protein UspA	57	62	182	475
AR505_0609	low molecular weight phosphotyrosine protein phosphatase	35	53	72	157
AR505_0749	trimethylamine permease MttP	47	27	119	1143
AR505_0750	hypothetical protein	33	27	72	955
AR505_0799	<i>O</i> -acetylserine sulfhydrylase CysM	20	15	13	43
AR505_0850	hypothetical protein	10	3	18	140
AR505_0859	transcriptional regulator TetR family	6152	7818	14364	31275
AR505_1008	SAM-dependent methyltransferases	16	42	34	89
AR505_1022	hypothetical protein	38	22	34	245
AR505_1066	bifunctional Methylthiol:corrinoid methyltransferase MtsA	52	63	105	881
AR505_1068	4Fe-4S ferredoxin iron-sulfur binding domain-containing protein	149	116	222	1139
AR505_1185	hydrogenase nickel insertion protein HypA1	14	15	34	1017
AR505_1242	dimethylamine permease MtbP1	29	17	58	223
AR505_1310	small multidrug resistance protein	30	11	34	163
AR505_1311	small multidrug resistance protein	29	5	30	99
AR505_1328	monomethylamine methyltransferase MtmB2	132	72	145	559
AR505_1329	methyltransferase cognate corrinoid protein MtmC1	75	49	76	291
AR505_1559	adhesin-like protein	35	28	48	418
Genes with 2-fold lower expression					
AR505_0183	hypothetical protein	43	115	63	21
AR505_0396	hypothetical protein	46	64	59	16
AR505_0466	HTH/CBS domain-containing protein	52	20	12	4
AR505_0471	phosphopyruvate hydratase Eno	48	55	28	21
AR505_1053	hypothetical protein	115	143	115	49
AR505_1219	transposase IS605 OrfB family	99	124	112	42
AR505_1622	F <sub>420</sub> H <sub>2</sub> dehydrogenase subunit N FpoN	538	210	705	86
AR505_1623	F <sub>420</sub> H <sub>2</sub> dehydrogenase subunit M FpoM	359	150	411	50
AR505_1624	F <sub>420</sub> H <sub>2</sub> dehydrogenase subunit L FpoL	541	232	589	86
AR505_1626	hypothetical protein	453	181	439	66
AR505_1628	F <sub>420</sub> H <sub>2</sub> dehydrogenase subunit I FpoI	633	214	546	82
AR505_1629	F <sub>420</sub> H <sub>2</sub> dehydrogenase subunit H FpoH	376	149	289	60
AR505_1630	F <sub>420</sub> H <sub>2</sub> dehydrogenase subunit D FpoD	721	317	468	134
AR505_1631	F <sub>420</sub> H <sub>2</sub> dehydrogenase subunit C FpoC	765	437	519	135
AR505_1632	F <sub>420</sub> H <sub>2</sub> dehydrogenase subunit B FpoB	373	230	261	69
AR505_1718	hypothetical protein	38	63	82	17
AR505_1759	ribosomal protein L7Ae Rpl7ae	466	268	221	50
AR505_1766	hypothetical protein	194	182	151	60



**Figure 5.27 Network diagram showing associations between methanogenesis genes and treatments based on Correspondence Analysis.** The threshold used was 0.1. Solid lines represents positive association, dotted lines represent negative associations. Line colours signify magnitude of association as illustrated by the key. Methanol:corrinoid methyltransferase (MtaB), methanol corrinoid protein (MtaC), monomethylamine:corrinoid methyltransferase (MtmB), monomethylamine corrinoid protein (MtmC), dimethylamine:corrinoid methyltransferase (MtbB), dimethylamine corrinoid protein (MtbC), dimethylamine permease (MtbP), trimethylamine permease (MttP), heterodisulfide reductase (Hdr). Treatment groups are defined in Table 5.3.

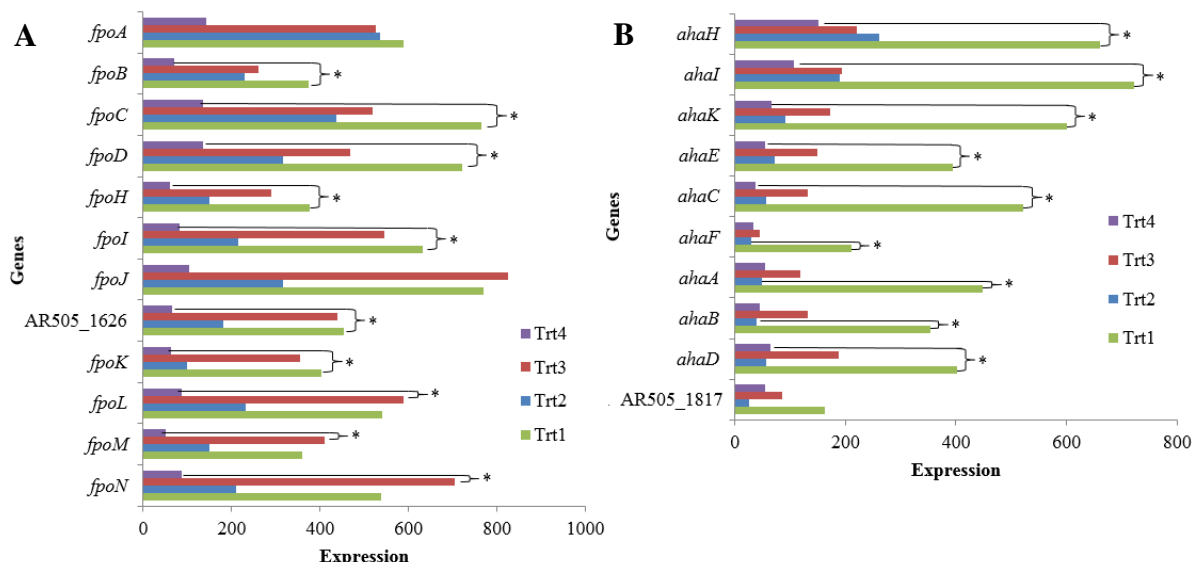
The association between methanogenesis genes and treatments used was examined using, Correspondence Analysis (Figure 5.27). Trt1 and Trt4 showed the greatest number of associations with methanogenesis genes, while only one gene was associated with Trt2 and two genes were associated with Trt3.

*R. flavefaciens* FD1 was used to provide H<sub>2</sub> continuously at a low concentration, therefore the influence of FD1 on the expression of ISO4-H5 was assessed by comparing genes showing two-fold difference in expression between Trt1 and the other three treatments (Table 5.8). Among the genes with higher expression in Trt1 were all nine subunits of the A<sub>1</sub>A<sub>0</sub> ATP synthase complex (Figure 5.27B), which were strongly up-regulated in the absence of *R. flavefaciens*.



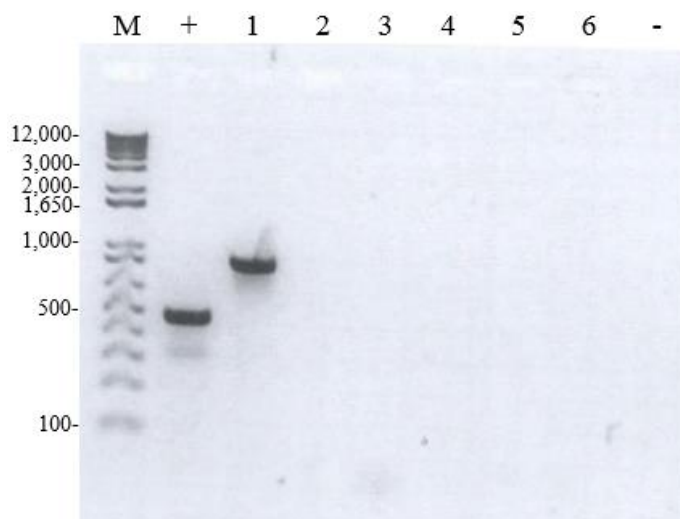
**Table 5.8 Genes with 2-fold expression difference in Trt1 compared with other**

Locus_tag		Product	Expression (RPK0.1M)			
			Trt1	Trt2	Trt3	Trt4
Genes with higher expression in Trt1						
AR505_0005	adhesin-like protein		462	152	153	76
AR505_0044	tRNA 2'-O-methylase		170	48	65	71
AR505_0093	molecular chaperone GrpE		309	134	151	100
AR505_0094	chaperone protein DnaK1		925	302	420	275
AR505_0114	methylase involved in ubiquinone/menaquinone biosynthesis		103	37	27	47
AR505_0201	Asp-tRNA <sup>Asn</sup> /Glu-tRNA <sup>Gln</sup> amidotransferase subunit B GatB		72	21	34	25
AR505_0287	sodium/proline symporter PutP		141	56	39	32
AR505_0418	ABC transporter permease protein		70	10	8	11
AR505_0466	HTH/CBS domain-containing protein		52	20	12	4
AR505_0594	transmembrane protein		326	57	41	26
AR505_0601	hydroxymethylglutaryl-CoA synthase		672	197	271	188
AR505_0602	acetyl-CoA acetyltransferase		1081	337	480	277
AR505_0603	DNA-binding protein		529	141	257	91
AR505_0605	ABC transporter permease protein		142	26	38	66
AR505_0897	hypothetical protein		166	72	60	33
AR505_0954	hypothetical protein		147	30	46	45
AR505_1016	hypothetical protein		134	41	21	30
AR505_1055	site-specific recombinase		97	13	10	29
AR505_1092	CRISPR-associated protein Cas7/Cse4/CasC		67	12	21	10
AR505_1177	MMPL domain-containing protein		69	15	10	19
AR505_1222	ferrous iron transport protein B FeoB		89	43	16	32
AR505_1263	molybdenum cofactor biosynthesis protein MoaA1		189	24	64	35
AR505_1266	hypothetical protein		283	53	35	27
AR505_1297	cobalamin 5'-phosphate synthase CobS		115	45	39	26
AR505_1369	ATP-dependent DNA helicase		73	19	31	32
AR505_1758	ribosomal protein S28e Rps28e		205	92	87	49
AR505_1818	A <sub>1</sub> A <sub>0</sub> ATP synthase subunit D		402	57	188	64
AR505_1819	A <sub>1</sub> A <sub>0</sub> ATP synthase subunit B		353	40	131	44
AR505_1820	A <sub>1</sub> A <sub>0</sub> ATP synthase subunit A		449	48	118	55
AR505_1821	A <sub>1</sub> A <sub>0</sub> ATP synthase subunit F		211	29	44	33
AR505_1822	A <sub>1</sub> A <sub>0</sub> ATP synthase subunit C		522	56	132	37
AR505_1823	A <sub>1</sub> A <sub>0</sub> ATP synthase subunit E		395	71	150	55
AR505_1824	A <sub>1</sub> A <sub>0</sub> ATP synthase subunit K		600	92	172	66
AR505_1825	A <sub>1</sub> A <sub>0</sub> ATP synthase subunit I		722	189	193	107
AR505_1826	A <sub>1</sub> A <sub>0</sub> ATP synthase subunit H		660	261	220	151
Genes with lower expression in Trt1						
AR505_0059	radical SAM domain protein		77	161	192	177
AR505_0377	NAD synthetase NadE		36	107	129	95
AR505_0619	transmembrane protein		11	660	349	727
AR505_0740	TATA-box-binding protein Tbp		68	258	140	207
AR505_0931	phage integrase		19	164	81	111
AR505_1008	SAM-dependent methyltransferase		16	42	34	89
AR505_1108	hypothetical protein		198	524	740	498
AR505_1157	hypothetical protein		4	48	27	23
AR505_1667	5-formaminoimidazole-4-carboxamide-1-(beta)-D- ribofuranosyl 5'-monophosphate-formate ligase PurP		116	460	365	302



**Figure 5.28 Differentially expressed ISO4-H5 genes during growth under low and high H<sub>2</sub> levels and different methyl sources. A.** Expression of genes encoding the Fpo-like complex. Fpo-like complex subunits (*fpo*). hypothetical transmembrane protein (ISO4H5\_1626). **B.** Expression of genes encoding A<sub>1</sub>A<sub>0</sub> ATP synthase subunits (*aha*). hypothetical protein immediate downstream of *ahaD* gene (ISO4H5\_1817). \*: Differentially expressed genes with FDR <0.05, KW p <0.05, fold change >2. Treatment groups are defined in Table 5.3.

During the analysis of gene expression across the ISO4-H5 genome, a region from AR505\_0313 to AR505\_0358 (estimated region 32.3 kb) was observed displayed no gene expression at all across all the treatments. This region was identified previously in Chapter 3 as being potentially horizontally transferred, and carries several elements consistent with an integrated prophage sequence. To assess whether this region contains a prophage sequence, PCR amplification was carried out on freshly isolated ISO4-H5 gDNA with primers (H5PHX.p1ca, H5PHX.q1ca) located outside the putative prophage region facing inwards, and with five primers located within the region. A product of only ~850 bp was obtained from the inward facing PCR, and all PCR amplifications within the putative prophage region failed to give a PCR product (Figure 5.29). This suggested that the putative prophage sequence was no longer present in the ISO4-H5 gDNA used for the PCR reactions. It is likely that this region had self-excised from the ISO4-H5 genome during the period of sub-culture between the original gDNA extraction for genome sequencing and this most recent gDNA extraction. This putative prophage region was therefore removed from the subsequent statistical assessments.



**Figure 5.29 PCR amplifications flanking and within the AR505\_0313 - AR505\_0358 region.** M: 1kb+ Marker, sizes are displayed in bp. +: Positive control 16S rRNA region amplified with primer pair 915af and 1386r. 1: PCR using primer pairs H5PHX.p1ca, H5PHX.q1ca flanking the AR505\_0313 – AR505\_0358 region. 2: PCR using primer pairs H5PH1.p1ca, H5PH1.q1ca within the AR505\_0313 – AR505\_0358 region. 3: PCR using primer pairs H5PH2.p1ca, H5PH2.q1ca within the AR505\_0313 – AR505\_0358 region. 4: PCR using primer pairs H5PH3.p1ca, H5PH3.q1ca within the AR505\_0313 – AR505\_0358 region. 5: PCR using primer pairs H5PH4.p1ca, H5PH4.q1ca within the AR505\_0313 – AR505\_0358 region. 6: PCR using primer pairs H5PH5.p1ca, H5PH5.q1ca within the AR505\_0313 – AR505\_0358 region. -: negative control PCR omitting genomic DNA.

### 5.2.7 Differential gene expressions in *S. dextrinosolvens* H5

A total of 85 *S. dextrinosolvens* H5 genes were differentially expressed between the four different treatments used in the ISO4-H5 co-culture experiments (Figure 5.21D, Table A.5.4). of which 74 are differentially expressed between Trt1 and Trt2, Trt1 and Trt3, Trt1 and Trt4, which is contributed by presence and absence of FD1, but only 11 genes were differentially expressed between Trt2, Trt3 and Trt4. Noteworthy genes with higher expression in absence of FD1 includes succinate dehydrogenase and fumarate reductase Fe-S protein (T508DRAFT\_01506), fumarate reductase (T508DRAFT\_01507) and malate dehydrogenase (T508DRAFT\_00914) involved in TCA cycle, pyruvate kinase (T508DRAFT\_01509) and phosphoglycerate mutase (T508DRAFT\_01623) involved in glycolysis, sugar phosphate permease (T508DRAFT\_00592),  $\alpha$ -1,4-glucan: $\alpha$ -1,4-glucan 6-glycosyltransferase (T508DRAFT\_01412) and glycogen/starch/ $\alpha$ -glucan phosphorylase (T508DRAFT\_00525) involved in carbohydrate utilisation, DNA polymerase III  $\epsilon$  subunit (T508DRAFT\_00876) involved in replication, ribosomal protein S1 (T508DRAFT\_01805), S9 (T508DRAFT\_00953), L18 (T508DRAFT\_00179), S11 (T508DRAFT\_00172), S13 (T508DRAFT\_00173), translation initiation factor IF-2 (T508DRAFT\_01537) and translation elongation factor TU (T508DRAFT\_02129, T508DRAFT\_02302) involved in translation, and conjugative transfer signal peptidase *traF* (T508DRAFT\_00891) involved in conjugative

transfer. Only five genes were observed with higher expression in presence of FD1, hypothetical proteins (T508DRAFT\_01024, T508DRAFT\_02014, T508DRAFT\_02218), predicted Zn-dependent proteases (T508DRAFT\_00450) and tRNA(Ile)-lysine synthetase (T508DRAFT\_00983).

#### 5.2.8 Differential gene expression in *R. flavefaciens* FD1

There were 378 genes differentially expressed in *R. flavefaciens* FD1 (Figure 5.22D; Table A.5.5), of which 375 were differentially expressed between Trt2 and Trt4 or between Trt3 and Trt4, but only 3 genes were differentially expressed between Trt2 and Trt3 (Table 5.9). As FD1 is not included in Trt1, there is no Trt1 data. Large numbers of genes were identified as differentially expressed due to the generally low number of transcripts (1% or less) mapped to *R. flavefaciens* FD1 genome in Trt4 (Table 5.5), which led to a high number of genes with no detectable expression in Trt4. A few genes in Trt4 still have a comparable expression, including a glycoside hydrolase family 43 (FD1\_0743), xylanase (FD1\_0753), type II secretory system (FD1\_0908 – FD1\_0914), cytochrome b/b6 (FD1\_1061), hypothetical proteins (FD1\_0919, FD1\_1202, FD1\_1472, FD1\_2347, FD1\_2848, FD1\_2992, FD1\_3943), aspartate—ammonia ligase (FD1\_1546), excisionase (FD1\_1976), resolvase (FD1\_1991), type 2 lantibiotic biosynthesis protein LanM (FD1\_2204, 2205), L-threonine ammonia-lyase (FD1\_2927), ribosome biogenesis GTP-binding protein YsxC (FD1\_3379), antitoxin HicB (FD1\_3635), RNA polymerase sigma factor (FD1\_3938).

**Table 5.9 Genes differentially expressed between Trt2 and Trt3**

Locus_tag	Product	Expression (RPK0.1M)		
		Trt2	Trt3	Trt4
<b>FD1_0077</b>	hypothetical protein	2461	91	282
<b>FD1_2086</b>	histidine kinase-DNA gyrase B-and HSP90-like ATPase	8	274	81
<b>FD1_3634</b>	serine/threonine protein kinase	200	3846	270

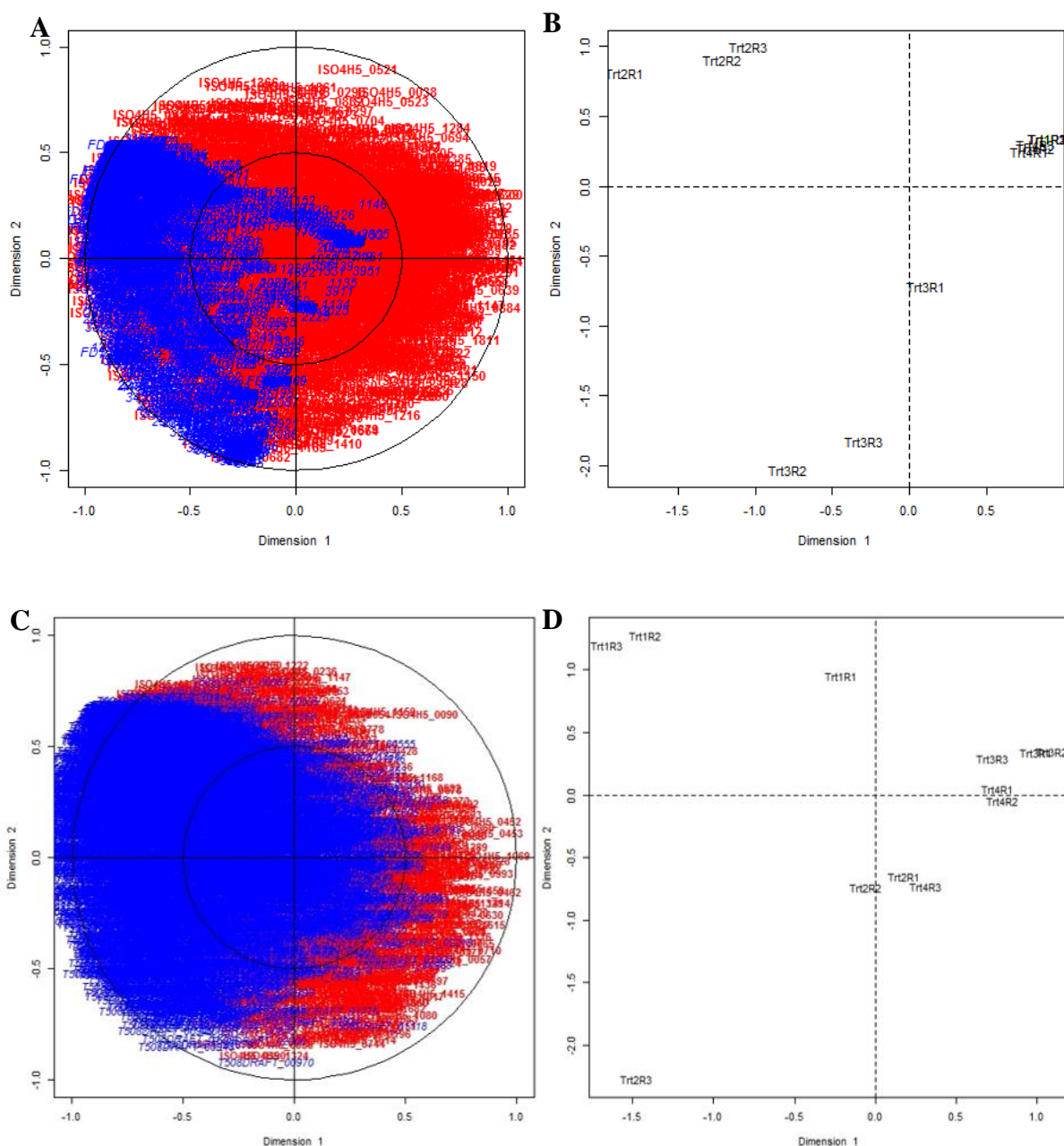
In order to assess the effect of H<sub>2</sub> level on differential gene expression without being biased by the low transcript levels in Trt4, the analysis was carried out between Trt2 and Trt3 only, which identified 464 genes differentially expressed with FDR < 0.05, fold change > 2 (Locus\_tag\*\* in Table A.5.5), but with a KW threshold of  $p > 0.05$ . Of these genes, which 382 had higher expression in Trt2, and 82 genes had higher expression in Trt3 (Table A.5.5). A large number of genes were highly expressed in Trt2, including carbohydrate utilisation (including cellulosome components), central carbon metabolism, nitrogen metabolism, electron transfer, transporters, cell surface proteins and secretion systems, cell replication and house-keeping, cofactor biosynthesis, and amino acid biosynthesis. Most noteworthy were the highly

expressed type II secretion system and type IV conjugative transfer system genes (Table A.5.6). A smaller range of genes were up-regulated in Trt3, including a region encoding lantibiotic biosynthesis (FD1\_2200 – FD1\_2209) along with a number of hypothetical proteins (FD1\_2191 – FD1\_2197) which appear to be part of this lantibiotic biosynthesis region (Table A.5.6). Other genes up-regulated in Trt3 encoded carbohydrate utilisation, cell surface proteins, a toxin/antitoxin system, transporters, secreted proteins, a DNA restriction/modification system and RNA polymerases.

#### 5.2.9 Associations between the ISO4-H5 and *R. flavefaciens* FD1, *S. dextrinosolvens* H5 transcriptomes

The associations between the ISO4-H5 and *R. flavefaciens* FD1 transcriptomes were explored by Regularised Canonical Correlation analysis (rCCA), which deals with the high dimensionality and low number of replicates of transcriptomics data by adding a regularization term on the diagonal of covariance matrices, but requires the number of dimensions to be tuned.

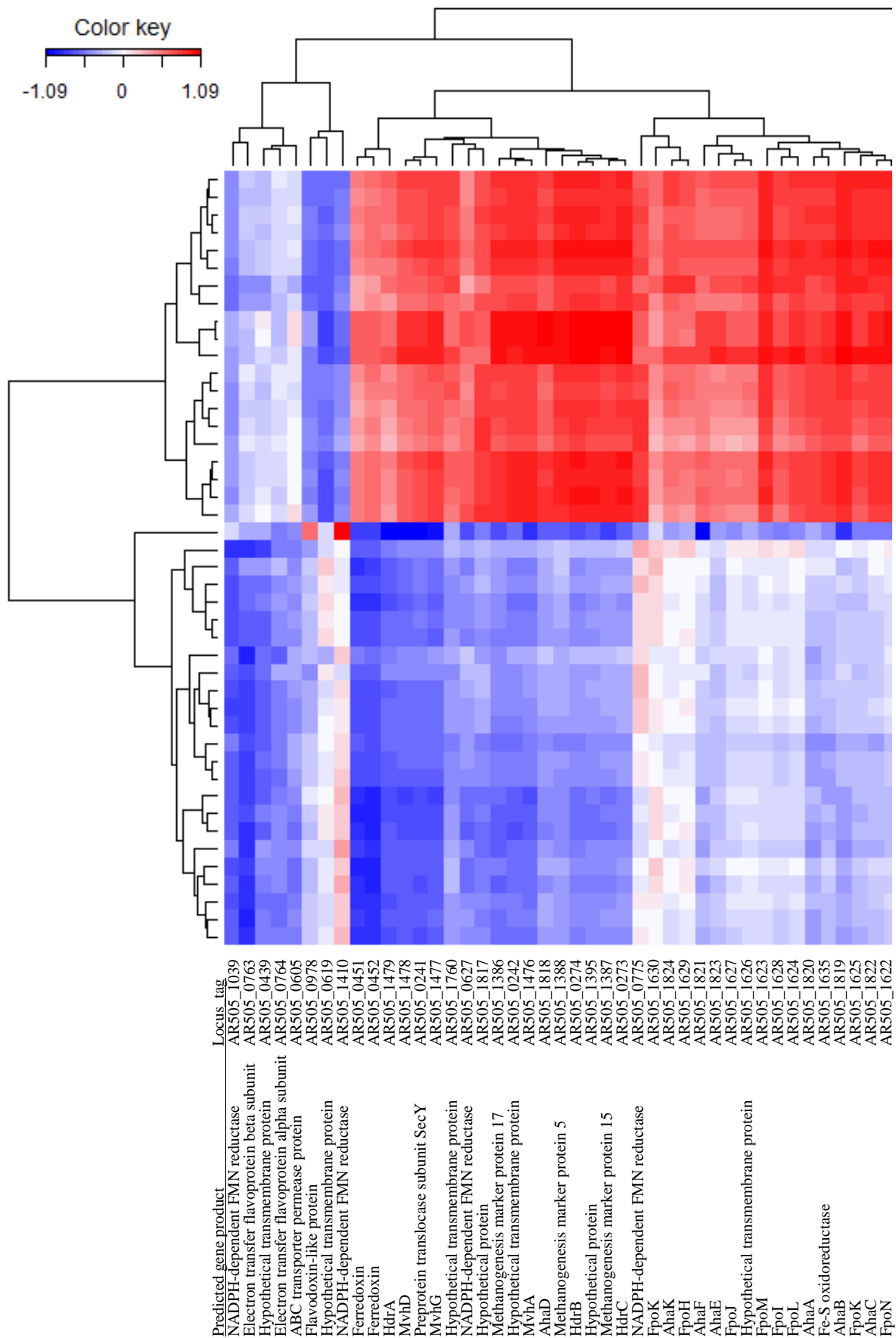
The *R. flavefaciens* FD1 gene expression variable shared coordinates with approximately one third of the ISO4-H5 gene expression variable, as seen by the overlap of coordinates to the left of the correlation circle plot (Figure 5.30A). This signifies that FD1 gene expressions have positive correlations with approximately one third of ISO4-H5 gene expressions, while negatively associated with the ISO4-H5 genes on the opposite side of the correlation circle plot. Based on the association between gene expression variables, the gene expressions in both genomes are negatively associated between Trt2 and Trt3 in dimension 2 (Figure 5.30B). The *S. dextrinosolvens* H5 gene expression variable shared coordinates with approximately two third of the ISO4-H5 gene expression variable, as seen by the overlap of coordinates in the middle and to the left of the correlation circle plot (Figure 5.30C). The overlap of gene coordinates to the left signifies a portion of genes in ISO4-H5 and *S. dextrinosolvens* H5 are positively associated, while the genes coordinates overlapped in the middle signifies these genes are neither positively or negatively associated with each other. Based on the association between gene expression variables, the gene expressions in both genomes are positively associated between Trt2 and Trt3 (Figure 5.30D).



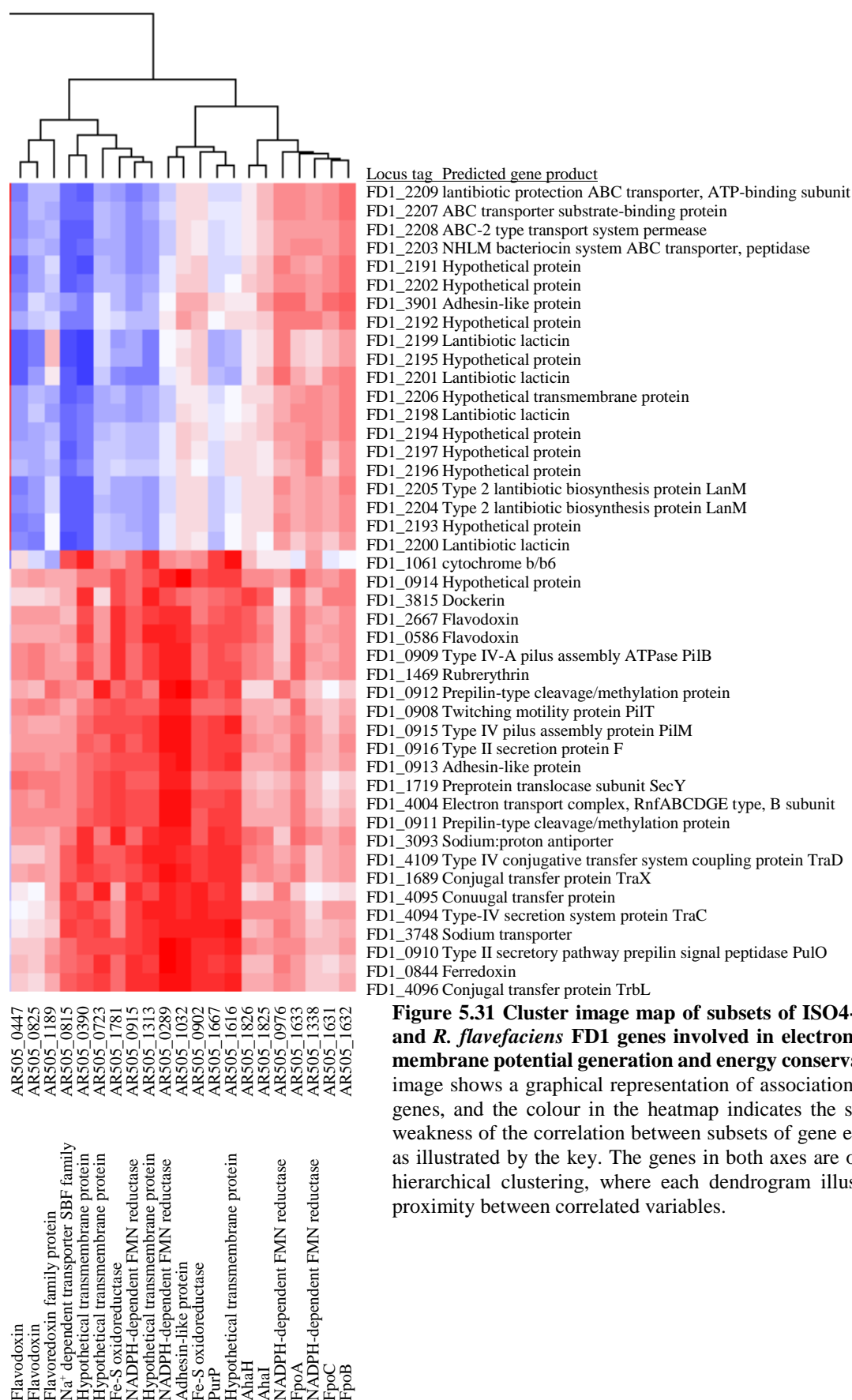
**Figure 5.30 Regularised Canonical Correlation analysis of associations between ISO4-H5 and *R. flavefaciens* FD1, *S. dextrinosolvens* H5 transcriptomes.** **A.** Correlation circle plot of ISO4-H5 and *R. flavefaciens* FD1 transcriptomes. Each gene vector represents variables, coordinate of vector represents correlation between variable and their associated component. ISO4-H5 genes are shown in red, *R. flavefaciens* FD1 genes are shown in blue. **B.** Correlation plot of treatments between ISO4-H5 and *R. flavefaciens* FD1. **C.** Correlation circle plot of ISO4-H5 and *S. dextrinosolvens* H5 transcriptomes. Each gene vector represents variables, coordinate of vector represents correlation between variable and their associated component. ISO4-H5 genes are shown in red, *S. dextrinosolvens* H5 genes are shown in blue. **D.** Correlation plot of treatments between ISO4-H5 and *S. dextrinosolvens* H5.

The finding that the expression of the ISO4-H5 A<sub>1</sub>A<sub>0</sub> ATP synthase and Fpo-like complex were negatively correlated with the presence of *R. flavefaciens* FD1 in the co-culture (Table A.5.3, Trt1 vs Trt2) was unexpected. Both the A<sub>1</sub>A<sub>0</sub> ATP synthase and Fpo-like complex are transmembrane enzyme complexes, involved in electron transfer, membrane potential

generation and energy conservation, therefore it was hypothesized that these functions in ISO4-H5 was influenced by the presence of *R. flavefaciens* FD1 in the co-culture. To explore this putative association, the ISO4-H5 and FD1 genes involved in electron transfer, membrane potential generation and energy conservation, that were differentially expressed between Trt2 and Trt3, were analysed using rCCA (using thresholds of  $\lambda_1 = 0.01099$ ,  $\lambda_2 = 0.11089$ ). *R. flavefaciens* FD1 genes also included transmembrane proteins involved in electron transfer or secretion, including type II secretion system, type IV conjugal transfer, lantibiotic biosynthesis and electron transfer. The gene associations gathered from rCCA are summarised in a clustered image map (Figure 5.31). The genes encoding a putative lantibiotic biosynthetic region in FD1 displayed strong positive associations with seven subunits of the ISO4-H5  $A_1A_0$  ATP synthase, two ferredoxins, *hdrABC*, *mvhAGD* genes, as well as genes of eight subunits of the Fpo-like complex. FD1 genes encoding cytochrome b/b6, type IV conjugal transfer system and type II secretion system displayed strong positive relationships with three hypothetical transmembrane proteins, one adhesin-like protein and two NADPH-dependent FMN reductase genes from ISO4-H5, and negative associations to the two ferredoxins.







### 5.3 Discussion

Methanomassiliicoccales is a recently established order of methanogens, and currently includes four members identified from human and termite gut sources (Borrel *et al.* 2012; Gorlas *et al.* 2012; Borrel *et al.* 2013; Lang *et al.* 2015). Only *Methanomassiliicoccus luminyensis* B10 has been isolated as a pure culture (Dridi *et al.* 2012). In this thesis, the methanogenic archaeon ISO4-H5 enriched from the ovine rumen has been isolated as a pure culture by exploiting the metabolism of *S. dextrinosolvens* H5 that was co-enriched with ISO4-H5. The purification of ISO4-H5 allowed confirmation that it is a small coccus, without fluorescence at 420 nm. Studies on human faeces and the termite gut also identified Methanomassiliicoccales as coccus-shaped archaea, however, controversy remains in the literature regarding their autofluorescence at 420 nm. Autofluorescence has been reported in the human-sourced *Methanomassiliicoccus luminyensis* (Dridi *et al.* 2012), while an absence of autofluorescence was reported in “*Candidatus* Methanoplasma termitum” from the termite gut (Lang *et al.* 2015). As discussed in Chapter 4, the analysis of available Methanomassiliicoccales genomes did not support their use of, or dependence on, F<sub>420</sub>.

The *S. dextrinosolvens* bacterium observed in the original ISO4-H5 enrichment, was also encountered in enrichments of two other members of Methanomassiliicoccales from the ovine rumen, ISO4-G1 and ISO4-G11. All three strains of *S. dextrinosolvens* co-enriched with rumen Methanomassiliicoccales, displayed 99% 16S rRNA sequence identity to *Succinivibrio dextrinosolvens* strain 0554 (NR026476) (Jeyanathan 2010). These observations suggest that *S. dextrinosolvens* may play a role in the supply of metabolites that support the growth of rumen Methanomassiliicoccales. A co-culture has also been observed between an anaerobic rumen fungus and Methanomassiliicoccales (Jin *et al.* 2014) suggesting that associations between Methanomassiliicoccales and other rumen microorganisms are common.

*S. dextrinosolvens* was described originally as a curved rod that fermented glucose, D-xylose, L-arabinose, maltose, galactose, fructose, sucrose, mannitol, dextrin and pectin and produced succinate, formate, acetate and lactate as the principal fermentation products (Bryant and Small, 1956). The medium used for ISO4-H5 enrichment contained approximately 38 mM acetate (Saleem *et al.* 2013), and 60 mM of formate. The additional supplementation of 10 mM succinate did not improve the growth of ISO4-H5 ruling out succinate as the stimulatory factor enhancing ISO4-H5 growth (Figure 5.19). Pectin addition to the ISO4-H5 enrichment culture

stimulated CH<sub>4</sub> production significantly ( $p = 5.6 \times 10^{-5}$ ), whereas the stimulation by glucose supplementation was less significant ( $p = 1.4 \times 10^{-2}$ ) (Figure. 5.01).

This led to the conclusion that pectin-grown *S. dextrinosolvens* produced a metabolite(s) stimulatory to ISO4-H5 growth, and the stimulation of growth using SSPGMS indicated that this metabolite(s) could be supplied without the need of the *S. dextrinosolvens* organism itself. The addition of the SSPGMS combined with ampicillin enabled isolation of ISO4-H5 from the enrichment culture. The purified culture however, did not grow as well as the enrichment culture, thus the SSPGMS supplementation could not fully replace *S. dextrinosolvens* (Figure 5.04). It is possible that certain stimulatory metabolites in the SSPGMS are unstable and degrade quickly once removed from *S. dextrinosolvens*, or possibly that physical interaction between ISO4-H5 and *S. dextrinosolvens* H5 is required to effect full stimulation of growth. However, physical contact between these two organisms has never been observed during microscopy of ISO4-H5 enrichment cultures.

In order to investigate the metabolites within SSPGMS that aided ISO4-H5 growth, NMR was carried out to identify compounds consumed during growth. Three compounds were identified as being significantly depleted after ISO4-H5 growth; methanol, ethanol and nicotinamide (Figure 5.07, 5.09, 5.10, 5.11). The methanol was likely used as a methyl substrate for methanogenesis, and was not limiting ISO4-H5 growth as it was supplied in the medium in excess of the amount used. Ethanol is likely utilized via an alcohol dehydrogenase (AR505\_0483), as described in Chapter 3, Section 3.2.4. Ethanol has been observed as a methyl source for methanogenesis in *Methanofollis ethanolicus* (Imachi *et al.* 2009), however, ISO4-H5 was unable to utilise ethanol alone as a methyl source (Figure 5.14A). Ethanol is also known to stimulate CH<sub>4</sub> formation in *M. ruminantium* M1 (Leahy *et al.* 2010), presumably via a coenzyme F<sub>420</sub>-dependent NADP reductase and NADP-dependent alcohol dehydrogenase (Berk and Thauer 1997). However, there is no evidence to support the utilisation of coenzyme F<sub>420</sub> in ISO4-H5, and no F<sub>420</sub>-dependent enzymes are predicted in ISO4-H5 genome, therefore ethanol is unlikely to supply reducing potential to methanogenesis in ISO4-H5 via this mechanism. Growth tests showed that ISO4-H5 could not utilise ethanol as a source of reducing potential in the absence of H<sub>2</sub> (Figure 5.16). Testing of 10 mM ethanol in the presence of H<sub>2</sub> did not influence the CH<sub>4</sub> production (Figure 5.17), which further supports the conclusion that ISO4-H5 does not utilise ethanol to supply reducing potential to methanogenesis. The alcohol dehydrogenase found in ISO4-H5 (AR505\_0483) shares homology with the multifunctional alcohol dehydrogenase from *Clostridium acetobutylicum*, which can produce butanol (Fischer

*et al.* 1993). Therefore, butanol was tested as a substrate, but it did not support CH<sub>4</sub> formation in ISO4-H5 (Figure 5.14A). An aldehyde dehydrogenase (AR505\_1599) was also predicted in ISO4-H5 genome, and together with an alcohol dehydrogenase could potentially produce acetyl-CoA from ethanol, and utilise ethanol as a carbon source. However, as indicated above, the addition of ethanol did not stimulate the growth rate or hydrogen utilisation of ISO4-H5 (Figure 5.17). The further assessment of alternative carbon sources was impractical as the medium that supports ISO4-H5 growth contains 30% rumen fluid, which provides 38 mM acetate (Saleem *et al.* 2013). Based on the genes present in the genome, acetate is likely the main carbon source of ISO4-H5. The assessment of alternative carbon sources in ISO4-H5 depends on the development a defined medium in which acetate can be replaced.

Nicotinamide was found to be present in SSPGMS at a level barely detectable by NMR, and it is likely utilised for NAD biosynthesis. However, ISO4-H5 possesses all the genes necessary to encode the production of NAD from aspartate or nicotinic acid (Chapter 3, Table 3.27). Nicotinic acid (0.1 mg/L) was also present in the cultivation medium as part of the vitamin mix, yet ISO4-H5 consumed additional nicotinamide from the SSPGMS. The growth of ISO4-H5 was not stimulated by additional nicotinamide or nicotinic acid, and does not appear to be limited by a lack of nicotinic acid (Figure 5.12). The use of exogenously supplied nicotinamide is likely to be preferable compared to *de novo* biosynthesis of NAD from aspartate, as it conserves energy, but it is not essential to the survival of ISO4-H5 and does not stimulate its growth.

The growth rate of ISO4-H5 culture *in vitro* is slow and it does not produce measurable turbidity at OD<sub>600</sub>. This is unlikely to be the case *in vivo*, otherwise ISO4-H5 would not survive being washed out of rumen due to the dilution effects of ingested feed and water. Therefore its *in vitro* growth is likely limited by some nutrient or growth factor. The amount of methanol and ethanol used during growth was quantified, and the amounts involved are not growth rate-limiting. It appears that while methanol, ethanol and nicotinamide supplied from SSPGMS may contribute to ISO4-H5 growth, the critical metabolite(s) from *S. dextrinosolvens* that stimulates growth is likely below the detection limit of NMR (~30 µM). If that is the case, then the critical metabolite probably functions as a cofactor for ISO4-H5 rather than a substrate. Fractionation of SSPGMS by HPLC and testing the ability of individual fractions, or combination of fractions, to stimulate ISO4-H5 growth, is the logical next step in identifying the metabolite(s) involved. While the identification of the limiting metabolite would be useful, the main

objective here was obtaining a pure culture of ISO4-H5 for subsequent experimentation, which was achieved.

The purification of ISO4-H5 allowed the inspection of its morphology as well as determining the substrates that support methanogenesis in this organism. The methanogen cell envelope has been found to be SDS- and protease-resistant (Kandler and König 1978), it consists of pseudomurein layer covered encrusted by S-layer, a two-dimensional protein crystal (Kandler and König 1978; Ellen *et al.* 2010). The pseudomurein is distinct to the bacterial peptidoglycan, as it lacks muramic acid and possess a different molar ratio of lysine:alanine:glutamate (Kandler and König 1978). The S-layer consisted of mostly glycoproteins containing *N*-linked glycans (Jarrell *et al.* 2010), in *Methanococcus voltae*, a 76 kDa protein in hexagonal lattice arrangement forms the S-layer (Koval and Jarrell 1987). The electron micrographies of *M. luminyensis* B10 did not show a prominent cell-wall-like structure (Dridi *et al.* 2012), furthermore, the cells were found to be SDS-sensitive (Dridi *et al.* 2012), suggesting the Methanomassiliicoccales possess a different cell envelope structure to other methanogens. ISO4-H5 appeared to have no cell wall and no observable extracellular matrix or capsule structure (Figure 5.03A). Instead, a bi-layer membrane was observed, with a thinner membrane (approximately 5 nm) compared to that previously observed in *M. luminyensis* B10 (approximately 100 nm, (Dridi *et al.* 2012)). ISO4-H5 also appears to be smaller (300 - 600 nm diameter) compared to *M. luminyensis* B10 (approximately 1200 nm diameter, (Dridi *et al.* 2012)). The lack of observed pseudomurein and S-layer can be reflected from genomic analysis in Chapter 3. Genomic study suggested synthesis of activated mannose and cell wall associated *N*-glycans are potentially possible (Borrel *et al.* 2014). As discussed in Chapter 3 and 4, the cell envelope may contain cardiolipin and glycosylated exopolysaccharide.

The ISO4-H5 genome sequence suggested ISO4-H5 can utilize methanol, mono-, di-, and tri-methylamines as well as certain forms of methylated thiols. Mono-, di-, and tri-methylamine (Figure 5.14), as well as methyl-3-methylthiopropionate (Figure 5.15) were confirmed as substrates for methanogenesis in ISO4-H5, in addition to methanol. The CH<sub>4</sub> production in ISO4-H5 cultures grown on trimethylamine was no different to that seen from growth on dimethylamine and methylamine, confirming that the methanogenic substrate is not the limiting condition for ISO4-H5 growth *in vitro*. Interestingly, ISO4-H5 was unable to grow on dimethylsulfide or methylmercaptopropionate, both were predicted as possible substrates based on the strong homology of its *mtsA* (AR505\_1066) gene to *Methanosarcina barkeri mtsA* gene (71.6% aa identity) (Tallant and Krzycki 1997). The chemical difference between

methylmercaptopropionate and methyl-3-methylthiopropionate is that the methyl-3-methylthiopropionate has a methyl group attached to the sulphur atom, which might make it more available for use by ISO4-H5. Interestingly, the dimethylsulfide has two methyl groups attached to the central sulfur atom, but the even charge distribution over the molecule may make it difficult for ISO4-H5 MtsA to attack. Betaine and choline can be broken down to trimethylamine, and methanogens have been observed to utilise betaine and choline directly (Muller *et al.* 1981; Fiebig and Gottschalk 1983; Wargo 2013; L'Haridon *et al.* 2014; Watkins *et al.* 2014). ISO4-H5 was unable to utilise choline and betaine (Figure 5.18), which suggests ISO4-H5 is dependent on other organisms within the rumen to supply methylamines, similar to what was found previously in *Methanosarcina barkeri* (Hippe *et al.* 1979).

The co-culture experiment between the ISO4-H5 enrichment culture and *R. flavefaciens* FD1 was carried out to identify changes in gene expression when ISO4-H5 was subjected to low or high H<sub>2</sub> conditions, and when growing on methylamine compared to methanol. *R. flavefaciens* FD1 was able to produce H<sub>2</sub> in excess of that required by ISO4-H5 (Figure 5.20C). The ISO4-H5 appears to have consumed more H<sub>2</sub> under the high hydrogen conditions (~1.2-1.5  $\mu$ mole, Figure 5.20B) compared to the low H<sub>2</sub> condition. However, CH<sub>4</sub> production in the low H<sub>2</sub> treatments was not much lower than that observed under high hydrogen. Assuming a 1:1 stoichiometry of H<sub>2</sub> use to CH<sub>4</sub> formation as predicted for ISO4-H5 growth on methanol, the FD1-containing co-cultures under low H<sub>2</sub> produced ~1  $\mu$ mole of H<sub>2</sub>, despite H<sub>2</sub> use appearing to be ~0.2-0.3  $\mu$ mole from the headspace analysis. Microbes that ferment carbohydrate and produce H<sub>2</sub> may suffer feedback inhibition in high H<sub>2</sub> environment (De Vrije and Claassen 2003). Similarly, *R. flavefaciens* FD1 may produce different level of H<sub>2</sub> in high H<sub>2</sub> and low H<sub>2</sub> environment due to feedback inhibition. . The production of CH<sub>4</sub> by ISO4-H5 appeared to occur sooner under the low H<sub>2</sub> condition (Figure 5.20A), which suggested there could be changes in gene expression that allows ISO4-H5 to scavenge H<sub>2</sub> earlier. The ISO4-H5 grown with the high H<sub>2</sub> condition produced more CH<sub>4</sub> at 120-hour than the low H<sub>2</sub> condition (Figure 5.20A), which suggests ISO4-H5 can achieve higher growth yield with a high H<sub>2</sub> condition. However, CH<sub>4</sub> production was observed earlier in both groups that grew under low H<sub>2</sub> condition, which suggests ISO4-H5 has higher growth rate in low H<sub>2</sub> condition. Within the rumen, a higher growth rate is beneficial for a micro-organism, as it allows the organism to grow more quickly and prevents it being washed out of the rumen. These observations suggest that ISO4-H5 associates with H<sub>2</sub> producing microorganisms to maintain a high growth rate, and therefore persist in the rumen.

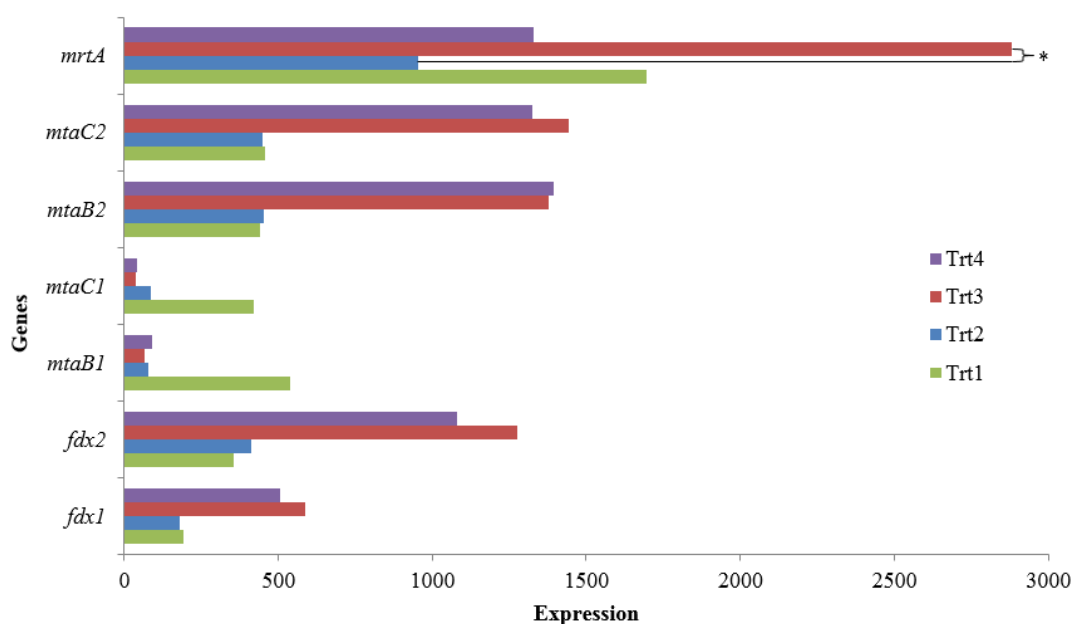
Transcriptomic studies requires the accurate interpretation of enormous amount of data, and multivariate statistical analysis is helpful in prioritizing differential gene expressions, and in filtering important and genuine relationships from the experimental noise and dependencies of the biological samples. The multivariate analysis of the gene expression from the co-culture studies showed clear differences between treatments in ISO4-H5. The most notable changes in gene expression between the high H<sub>2</sub> Trt2 and the low H<sub>2</sub> Trt3 were the expression of *hdr* and *mvh* that encode enzymes involved in the supply of reducing potential to methanogenesis using H<sub>2</sub> (Figure 5.24). The HdrA protein subunit harbors flavin adenine dinucleotide (FAD) necessary for reduction of CoM-S-S-CoB (Kaster *et al.* 2011), and both HdrA, HdrC also contain iron-sulfur cluster binding motifs. The HdrB subunit has 10 highly conserved cysteine motifs important in re-oxidation of CoM-S-S-CoB, while HdrD contains characteristics of both HdrB and HdrC, having the iron sulphur cluster binding motifs as well as two conserved cysteine motifs (Hedderich *et al.* 2005). The *hdrA*, *hdrB1*, *hdrC* and *mvhAGD* genes were up-regulated under low H<sub>2</sub> condition which suggests that ISO4-H5 up-regulates the expression of the corresponding Hdr and Mvh enzymes to scavenge additional H<sub>2</sub> under low H<sub>2</sub> conditions. The *hdrB1* and *hdrC* genes were identified as a conserved operon in all Methanomassiliicoccales genomes analysed (Appendix, Table 4.S.11), while *hdrB2* is not conserved, and display a significantly lower score in BLASTp, Pfam and COG analyses. This suggests *hdrB2* may not encode a genuine HdrB, but rather an iron-sulfur motif containing protein with unknown function. Conversely, the *hdrD* gene was not differentially expressed under different H<sub>2</sub> condition, which suggests HdrD does not form complex with HdrABC in ISO4-H5, or its function does not require up-regulation, it has been proposed to be coupled with the Fpo-like complex and function similarly to the absent FpoF and FpoO subunit (Lang *et al.* 2015), as discussed in Chapter 4.

The genes encoding MrtA, a hypothetical protein AR505\_1395 downstream of the *mrtA* gene and three methanogenesis marker proteins Mmp7, 15 and 17 were also found to be differentially expressed between Trt2 and Trt3. The up-regulation of the *mrtA* gene appears to have contributed to ISO4-H5 producing CH<sub>4</sub> earlier under the low H<sub>2</sub> conditions. The Mrt complex has been structurally characterised in *Methanosarcina barkeri* as a heterohexamer of  $\alpha\alpha'\beta\beta'\gamma\gamma'$  (Grabarse *et al.* 2000), therefore it is unexpected that similar up-regulation was not observed from the *mrtB* and *mrtG* genes. Similarly, *mrtA* specific up-regulation has also been observed previously in cows, which was suppressed by rapeseed oil-supplemented diet (Poulsen *et al.* 2013). The up-regulation of AR505\_1395 suggests this hypothetical protein is

inside the Mrt operon and regulated similarly to the *mrtA*, which signifies this protein could be associated with the MrtA. A recent rumen metagenome and metatranscriptome study has classified the *mcr/mrt* operons into three major groups (Shi *et al.* 2014), and Methanomassiliicoccales *mrt* operon is grouped differently to the *mcr* and *mrt* operons of *Methanobrevibacter*s (Shi *et al.* 2014). The classification combined with the differential expression of *mrtA* and AR505\_1395 genes suggests the Methanomassiliicoccales Mrt complex could behave differently to the Mrt complexes previously studied, and deserve further investigations in both its structure and function.

The genes encoding methanol utilization, and the ferredoxins *fdx1* (AR505\_0452) and *fdx2* (AR505\_0453) were not differentially expressed under differing hydrogen conditions according to the statistical cut off, however the patterns of gene expression displayed by *mtaC2*, *mtaB2* and ferredoxins ISO4H5\_0452 and ISO4H5\_0453 appeared to be strongly influenced by H<sub>2</sub> level. Most gene expression is controlled by multiple factors, for example *mrtA* (Figure 5.28), was up-regulated by low H<sub>2</sub> concentration but it was not up-regulated when methylamine is present. The *mtaB2* and *mtaC2* appear to be influenced by low H<sub>2</sub> condition, while the *mtaB1* and *mtaC1* appear to be unaffected by H<sub>2</sub> level, and seem to respond negatively to the presence of *R. flavefaciens* FD1. In *Methanosaeta* spp. which lack the *fpoF* gene of the Fpo-like complex, it has been proposed that ferredoxin can act as an electron carrier (Welte and Deppenmeier 2011). Similarly, all the Methanomassiliicoccales genomes analysed lack the *fpoF* gene (Chapter 4), and may also utilise ferredoxin for electron transport. The expression pattern of ferredoxins AR505\_0452 and AR505\_0453 most closely match those of *mtaB2*, *mtaC2*, *mrtA* (Figure 5.32), *hdrABC* and *mvhAGD* genes (Figure 5.25), and this similarity of expression pattern suggests these ferredoxins could be associated with Mvh, Hdr and may be directly involved in electron transfer for methanogenesis.





**Figure 5.32 Expression of genes involved in methanol utilizing methanogenesis between high and low H<sub>2</sub>.** *mtaB*: methanol:corrinoid methyltransferase. *mtaC*: methanol corrinoid protein. *mrt*: methyl-coenzyme M reductase complex. *fdx*: ferredoxin.

The importance of expression pattern in the co-culture study can again be seen when assessing methylamine-influenced gene expressions. Ideally, Trt3 and Trt4 should be compared as the methyl source is the only variable, however, comparison between Trt3 and Trt4 showed only 13 genes were differentially expressed, and none of the genes were directly involved in methyl substrate utilisation. When overall gene expression was analysed in a correspondence analysis, the strongest associations were linked to Trt1 and Trt4 (Figure 5.23). This is probably due to these two conditions being the most different, in terms of substrate type and H<sub>2</sub> level, and when genes are regulated by multiple factors, the differences between Trt1 and Trt4 would be greater than differences between Trt3 and Trt4. The genes proposed to be involved in non-methanol methyl substrate utilisation and their gene expressions were summarised (Figure 5.25) and the *mtmB2*, *mtmC1*, *mtbC*, *mtbB*, *mtbP1*, *mttP*, AR505\_1068 and *mtsA* genes were identified as being differentially expressed, either between Trt1 and Trt4 or between Trt2 and Trt4. All of these genes show a pattern of high gene expression in Trt4, even *mtsB* gene which was not identified as differentially expressed. Although the trimethylamine utilising *mttBC* genes were not differentially expressed it appears to be highly expressed across all treatments. The up-regulation of methyl-3-methylthiopropionate utilising genes *mtsAB* is unexpected, as there is no methyl-3-methylthiopropionate supplementation in this experiment. This suggests that in the rumen the availability of methylamines for ISO4-H5 is linked to the availability of methyl-

3-methylthiopropionate, and possibly other utilisable methylated thiol compounds, and falls under the same regulation mechanism. The elevated gene expression of AR505\_1068 encoding a 4Fe-4S ferredoxin iron-sulfur binding domain containing protein downstream of the *mtsB* gene showed it is potentially associated with methyl-3-methylthiopropionate utilisation.

Although Trt4 was intended to test methylamine alone as a methyl source for methanogenesis, in hindsight, approximately 2 mM of methanol would have been carried over in the inoculum in this experiment. Therefore Trt4 contained both methylamine (10 mM) and methanol (~2 mM) as methyl sources. Despite this, the differences in gene expressions between Trt3 and Trt4 were still obvious in the MDS, PCoA and group dispersion plot (Figure 5.21) analyses, which signifies that ISO4-H5 gene expression responded to the presence of methylamine despite the presence of methanol.

In order to identify genes that are influenced by the presence of methylamine and methanol, genes with two-fold or more expression difference between Trt4 and all other treatments were selected amongst the differentially expressed genes (Table 5.06). This identified some genes that could be highly important. This included the AR505\_1559 gene encoding an adhesin-like protein, which was approximately 10 fold higher expressed in Trt4 compared to all three other treatments, and it may be involved in adherence to a methylamine/methyl-3-methylthiopropionate producing bacteria within the rumen. The hypothetical protein AR505\_0750 adjacent to the *mttP* gene also showed at least 10 fold higher expression in Trt4 compared to other treatments, and may be associated with trimethylamine utilisation. Based on the high up-regulation of gene expression involved in methylamine utilisation in the presence of methanol and methylamine, methylamines could be the preferred substrate of ISO4-H5. The growth of ISO4-H5 on different methyl substrates appeared to be similar, however these growth tests were conducted under sub-optimal conditions as the growth limiting component of the medium remained unidentified. Genes encoding eight subunits of the Fpo-like complex have been identified as differentially expressed with two to seven fold lower expression in Trt4, this finding is unexpected, as theoretically the methylamines and methyl-3-methylthiopropionate shares methanogenesis genes with methanol.

Although the multivariate analysis showed clear differences between treatments in ISO4-H5, the differences were less significant in *S. dextrinosolvens* H5 (Figure 5.22), and smaller number of genes are differentially expressed under the conditions tested (Figure 5.22D). On the contrary, large number of genes were found to be differentially expressed in *R. flavefaciens*

FD1, although all but three are due to comparison to Trt4 (Figure 5.23D), which is likely due to the relatively low number of transcripts aligned to *R. flavefaciens* FD1 (Table 5.05B). Therefore the differential expressed genes identified in this manner between Trt4 and other treatments have questionable reliability.

*R. flavefaciens* FD1 was used to create a low H<sub>2</sub> environment in the co-culture with ISO4-H5. However, the *R. flavefaciens* FD1 in Trt4 has an abnormally low transcript abundance in comparison to Trt3, the low transcript abundance may not necessarily mean the lack of growth in *R. flavefaciens* FD1, as it was still able to supply sufficient quantity of H<sub>2</sub> to support the growth of ISO4-H5 (Figure 5.20C). The only variable between Trt3 and Trt4 is the presence of monomethylamine. Methylamine is considered a permeable weak base (Ohkuma and Poole 1978; Ritchie and Gibson 1987), at 10 mM it inhibited *R. flavefaciens* by approximately 20% (Ricke *et al.* 1994), this may explain the abnormal transcript abundance observed in Trt4. In a rumen perspective, this suggests the presence of methylamine-consuming methylotrophic methanogens would benefit *R. flavefaciens* FD1-like organisms by uptake of methylamine. This may also explain why methylamine was predicted as the conserved substrate across all the Methanomassiliicoccales genomes analysed, as growth on methylamine would keep methylamine levels low in the rumen and encourage the growth of *R. flavefaciens*-like organisms and ensure a steady source of H<sub>2</sub> for growth of Methanomassiliicoccales.

The presence of *R. flavefaciens* FD1 seems to influence ISO4-H5 gene expression in an unexpected way. The most significant finding was that the ISO4-H5 A<sub>1</sub>A<sub>0</sub> ATP synthase complex was highly expressed in Trt1 under high H<sub>2</sub> in the absence of FD1 (Table 5.07B, Figure 5.27B). The proton translocating F<sub>1</sub>F<sub>0</sub> ATP synthase of *Lactobacillus acidophilus* has been observed to be up-regulated at low pH (Kullen and Klaenhammer 1999). However, no difference in pH was detected between each of the treatments, therefore the change of expression in ATP synthase could probably be associated with *R. flavefaciens* FD1. The associations between transcriptomes of ISO4-H5 and *R. flavefaciens* FD1 (Figure 5.28) showed positive correlation between six A<sub>1</sub>A<sub>0</sub> ATP synthase subunits and Fpo-like complex of ISO4-H5 with a region in *R. flavefaciens* FD1 encoding a putative lantibiotic biosynthetic island (Figure 5.31). Lantibiotics are antimicrobial peptides with post-translational modification involving lanthionine or β-methyllanthionine (O'Sullivan *et al.* 2011). There are two types of lantibiotics, Type 1 and Type 2 (Begley *et al.* 2009). Type 1 lantibiotics are modified by LanBC, (e.g. nisin), and are involved in pore formation in cell membranes via binding to a cell wall precursor, lipid II (Hsu *et al.* 2004). Type 2 lantibiotics are modified by LanM (Begley *et*

al. 2009 7882), (e.g. lactacin 3147 which is in fact two peptides LtnA1 and LtnA2 that act synergistically), and also bind to lipid II and induce pore formation, resulting in ion leakage (Wiedemann *et al.* 2006). The lantibiotic biosynthetic island of *R. flavefaciens* FD1 genome is predicted to contain two LanM homologues FlvM1 and FlvM2, which suggests it produces Type 2 lantibiotics. Another species of the genus *Ruminococcus*, *R. gnavus* E1, could produce a Type 2A lantibiotic called ruminococcin A, which acts against bacteria phylogenetically related to *R. gnavus*, and *Clostridium* spp. (Dabard *et al.* 2001). A recently study expressed the FD1 lantibiotic operon in *E. coli* and verified that FD1 indeed produces Type 2 lantibiotics, including four structurally conserved lipid II binding peptides (Flv $\alpha$ .a-d) and eight structurally diverse  $\beta$ -peptides (Flv $\beta$ .a-g), which resembles the two-component lantibiotic system such as lactacin 3147, in particular, FlvM1 modified Flv $\alpha$ .a and Flv $\beta$ .g displayed synergistic antimicrobial effect against *Micrococcus luteus* DSM 1790 (Zhao and van der Donk 2016). The FD1 did not produce detectable level of lantibiotics when co-cultured with *Ruminococcus albus* 7 (Zhao and van der Donk 2016), which suggested that Flv $\alpha$ s and Flv $\beta$ s may behave differently to ruminococcin A, it may not target phylogenetically related bacteria, or there are additional triggers of expression. This means that ISO4-H5 is may be affected by the lantibiotics, which is reflected in the up-regulation of Fpo-like complex and ATP synthase, which suggests the proton gradient of ISO4-H5 was affected. This study offers no conclusive evidence on this matter, as the cell envelope structure of ISO4-H5 is unknown.

Negative correlations were also identified between the *R. flavefaciens* FD1 cytochrome b/b6, type II secretion and type IV conjugative transfer systems genes and the ISO4-H5 A<sub>1</sub>A<sub>0</sub> ATP synthase subunits, Mvh, Hdr and ferredoxins up-regulated under the low H<sub>2</sub> condition. *Geobacter sulfurreducens* has been reported as being able to transfer electrons to extracellular electron acceptors via pili (Reguera *et al.* 2005), and direct interspecies electron transfer has been observed between *Geobacter metallireducens* and two methanogens, *Methanosaeta harundinacea* (Rotaru *et al.* 2014) and *Methanosarcina barkeri* (Rotaru *et al.* 2014). Furthermore, co-cultures with these methanogens could not be established with a pilin-deficient *G. metallireducens* strain (Rotaru *et al.* 2014), which suggests the electron transfer via electrical conducting pili is an important form of syntrophy. Additionally, intercellular “wiring” between sulphate-reducing bacteria and anaerobic methanotrophic archaea was found to be essential for the anaerobic oxidation of CH<sub>4</sub> (Wegener *et al.* 2015). In the co-culture of *G. metallireducens* and *G. sulfurreducens*, the up-regulation of PilA, pilin domain 2 protein and pilus-associated cytochrome OmcS were considered to be important in the direct

interspecies electron transfer (Shrestha *et al.* 2013). The cytochrome b/b6 gene and genes involved in Type II secretion and Type IV conjugative transfer were up-regulated in *R. flavefaciens* FD1 during the high H<sub>2</sub> condition, which suggests *R. flavefaciens* FD1 may interact in a similar fashion with ISO4-H5, transferring electrons via pili and/or cytochromes. Up-regulation of these genes in FD1 may facilitate electron transport to ISO4-H5, disposing of reducing equivalents and alleviating the feedback inhibition caused by H<sub>2</sub> under high H<sub>2</sub> conditions. The ISO4-H5 A<sub>1</sub>A<sub>0</sub> ATP synthase and Fpo complex genes were down-regulated in Trt2 in the presence of *R. flavefaciens* FD1, which suggests ISO4-H5 may utilise electrons transferred from *R. flavefaciens* FD1 to drive proton efflux and ATP production. The up-regulation of the Fpo complex and A<sub>1</sub>A<sub>0</sub> ATP synthase in ISO4-H5 in the absence of *R. flavefaciens* FD1 may reflect its need to create its own membrane potential/proton gradient to drive ATP synthesis.

The majority of changes in the transcriptome of *S. dextrinosolvens* H5 was due to the presence of *R. flavefaciens* FD1. Most of the differentially expressed genes were up-regulated in absence of FD1, and are involved in carbohydrate utilisation, central carbon metabolism, protein biosynthesis and other house-keeping functions, which suggests that *S. dextrinosolvens* H5 may have compete with FD1 for utilisable carbohydrate. Although the rCCA suggested there are gene associations between ISO4-H5 and *S. dextrinosolvens* H5, because all four treatments contain both ISO4-H5 and *S. dextrinosolvens* H5, it is difficult to identify the particular interaction between them. The logical next step would be to sequence the transcriptome of *S. dextrinosolvens* H5 monoculture and identify the difference in gene expression, which may give hints to the interaction between *S. dextrinosolvens* H5 and ISO4-H5.

When comparing *R. flavefaciens* FD1 gene expression between Trt2 and Trt3, a large number of genes involved in cell replication and carbohydrate utilisation appeared to be highly expressed under the high H<sub>2</sub> condition (Table A.5.05), which suggests FD1 grew better at high H<sub>2</sub> condition. This is supported by the generally higher number of transcripts assigned to *R. flavefaciens* FD1 in Trt2 (Table 5.05B). However, this observation contradicts the dogma that high H<sub>2</sub> concentrations inhibit fibre-degrading rumen microbes, such as *R. flavefaciens* (Iannotti *et al.* 1973; Wolin 1976; Rees *et al.* 1995; Morvan *et al.* 1996). The interspecies H<sub>2</sub> transfer has been studied using FD1 and *Mbb. ruminantium* PS, when interspecies H<sub>2</sub> transfer happen between FD1 and methanogens, the molar yield of acetate in hexose fermentation should double, and the succinate production that was used to regenerate NAD<sup>+</sup> in monoculture should be reduced (Wolin *et al.* 1997). In this study, the pyruvate flavodoxin oxidoreductase

(FD1\_0183) and acetate kinase (FD1\_0693) genes involved in acetate and H<sub>2</sub> production were up-regulated in Trt2, as are six genes involved in glycolysis, phosphoglycerate kinase (FD1\_4123), triosephosphate isomerase (FD1\_0338), fructose-bisphosphate aldolase (FD1\_0313), 6-phosphofructokinase (FD1\_2874) and glucokinase (FD1\_0252). The elevated expression of genes involved in glycolysis in Trt2 contradicts the dogma that high H<sub>2</sub> concentrations inhibit fermentation. The growth of ISO4-H5 in low hydrogen condition is evident of the interspecies H<sub>2</sub> transfer, which is reflected in the expression of genes involved in acetate and H<sub>2</sub> production, however, the expression of these genes under high H<sub>2</sub> condition also suggests there could be an alternative method for FD1 to remove excess electrons, possibly by transferring the excess electrons to ISO4-H5. The succinate production pathway was used by FD1 to remove excess electrons in monoculture (Wolin *et al.* 1997), yet malate dehydrogenase (FD1\_3067), fumarate hydratase (FD1\_3371) and fumarate reductase (FD1\_3362) involved in succinate production were up-regulated in Trt2 instead of down-regulated, which is inconsistent with the previous findings. It is possible that the electron transfer from *R. flavefaciens* FD1 to ISO4-H5 postulated above, may relieve the feedback inhibition caused by high H<sub>2</sub> condition, by dumping electrons from NADH and allowing NAD to be recycled. Only a few FD1 genes were up-regulated in low hydrogen condition (Table 5.11B), most notable being the genes within the putative lantibiotic biosynthesis island. Which means *R. flavefaciens* FD1 may gain a competitive advantage over other rumen bacteria due to the production of antimicrobial lantibiotics.

In conclusion, ISO4-H5 carries out hydrogen dependent methylotrophic methanogenesis using methanol, mono-, di-, tri- methylamine and methyl-3-methylthiopropionate as substrate. Two sets of genes are involved in methanol utilisation, one set is up-regulated by low H<sub>2</sub> conditions. The methylamine permeases and the genes involved in utilising mono- and di-methylamine are up-regulated by presence of methylamine, as is the methyl-3-methylthiopropionate utilising operon *mtsAB* and ISO4H5\_1075 genes. H<sub>2</sub> level response has been observed in *mcrA*, *hdr* and *mvh* genes, two ferredoxins have been identified as associated with Hdr and Mvh. ISO4-H5 also appears to consume ethanol, but it is not used as a methyl substrate or as a source of reducing potential for methanogenesis.

## Chapter 6

### **The complete genome sequence of *Methanobrevibacter* sp. D5 and its comparison to other members of the genus *Methanobrevibacter***

#### **6.1 Introduction**

The genus *Methanobrevibacter* (*Mbb.*) consists of obligate anaerobes that are able to reduce carbon dioxide (CO<sub>2</sub>) to methane (CH<sub>4</sub>) using hydrogen (H<sub>2</sub>) and sometimes formate as the electron donor (Miller 2001). *Methanobrevibacter* spp. cannot utilise acetate, methanol, methylamines or other organic compounds as electron donors for methanogenesis (Miller 2001). Ammonium ions are considered the major source of cell nitrogen and acetate is expected to be the major source of cell carbon (Miller 2001). Typically, additional nutrients such as casamino acid, yeast extract or B-complex vitamins are required for good growth (Miller and Lin 2002). Cells are usually short, lancet shaped cocci, short rods to long, filamentous rods that are non-sporulating and non-motile (Balch *et al.* 1979). Members of the *Methanobrevibacter* species contain pseudomurein cell walls which are composed of *N*-acetyl amino sugars, neutral sugars and peptide moieties containing L-glutamate, L-alanine, L-lysine, L-ornithine (Kandler and König 1978). To date, 15 *Methanobrevibacter* type strains has been identified, and their environments and nutritional requirements are summarised in Table 6.1.

**Table 6.1 *Methanobrevibacter* species and their nutritional requirements**

Species <sup>1</sup>	Environments	Substrates	Growth requirements	Reference
<i>Methanobrevibacter smithii</i> <b>PS<sup>T</sup></b> , <b>TS145A</b> , <b>TS145B</b> , <b>TS146A</b> , <b>TS146B</b> , <b>TS146C</b> , <b>TS146D</b> , <b>TS146E</b> , <b>TS147A</b> , <b>TS147B</b> , <b>TS147C</b> , <b>TS94B</b> , <b>TS94C</b> , <b>TS95A</b> , <b>TS95B</b> , <b>TS95C</b> , <b>TS95D</b> , <b>TS96A</b> , <b>TS96B</b> , <b>TS96C</b> , GMS-01, B-181, ALI, F1	Sewage, human colon, human faeces, rumen	CO <sub>2</sub> , H <sub>2</sub> , formate	Acetate, NH <sub>4</sub> , B vitamins	Balch <i>et al.</i> , 1979, Bryant <i>et al.</i> , 1971, Miller <i>et al.</i> , 1982
<i>Methanobrevibacter ruminantium</i> <b>M1<sup>T</sup></b>	Bovine rumen	CO <sub>2</sub> , H <sub>2</sub> , formate	Acetate, CoM, NH <sub>4</sub> , AA, 2-methylbutyric acid, YE, TP	Smith and Hungate, 1958, Bryant <i>et al.</i> , 1971
<i>Methanobrevibacter wolinii</i> <b>SH<sup>T</sup></b>	Sheep faeces, rumen	CO <sub>2</sub> , H <sub>2</sub>	Acetate, YE, CoM, VFA	Miller and Lin, 2002
<i>Methanobrevibacter thaueri</i> <b>CW<sup>T</sup></b>	Cow faeces, rumen	CO <sub>2</sub> , H <sub>2</sub>	Acetate, TP, YE	Miller and Lin, 2002
<i>Methanobrevibacter woesei</i> <b>GS<sup>T</sup></b>	Goose faeces	CO <sub>2</sub> , H <sub>2</sub> , formate	Acetate, NaCl, TP, YE	Miller and Lin, 2002
<i>Methanobrevibacter gottschalkii</i> <b>HO<sup>T</sup></b> , PG	Horse faeces, pig faeces, rumen	CO <sub>2</sub> , H <sub>2</sub>	Acetate, NaCl, TP, YE	Miller and Lin, 2002
<i>Methanobrevibacter oralis</i> <b>ZR<sup>T</sup></b> , <b>JMR01</b>	Human subgingival plaque, human faeces	CO <sub>2</sub> , H <sub>2</sub>	Fecal extract, VFA	Ferrari <i>et al.</i> , 1994
<i>Methanobrevibacter olleyae</i> <b>KM1H5-1P<sup>T</sup></b> , <b>YLM1</b>	Ovine rumen	CO <sub>2</sub> , H <sub>2</sub> , formate	Acetate	Rea <i>et al.</i> , 2007
<i>Methanobrevibacter millerae</i> <b>ZA-10<sup>T</sup></b> , <b>SM9</b> , HW02, <b>YE315</b>	Bovine rumen	CO <sub>2</sub> , H <sub>2</sub> , formate	YE, TP	Rea <i>et al.</i> , 2007
<i>Methanobrevibacter boviskoreani</i> <b>JH1<sup>T</sup></b> , JH4, JH8, <b>AbM4</b>	Bovine rumen	CO <sub>2</sub> , H <sub>2</sub> , formate	YE, CoM, VFA	Lee <i>et al.</i> , 2013
<i>Methanobrevibacter arboriphilus</i> <b>DH1<sup>T</sup></b> , AZ, DC, <b>ANOR1</b> , A2, SH4, <b>SA</b>	Decaying cottonwood, paddy field soil, human faeces	CO <sub>2</sub> , H <sub>2</sub> , formate <sup>v</sup>	RF, YE, TP, NaCl (v), B vitamins	Zeikus and Henning, 1975, Miller 2001
<i>Methanobrevibacter cuticularis</i> <b>RFM-1<sup>T</sup></b>	Termite gut	CO <sub>2</sub> , H <sub>2</sub> , formate	RF, YE, CA	Leadbetter <i>et al.</i> , 1996
<i>Methanobrevibacter curvatus</i> <b>RFM-2<sup>T</sup></b>	Termite gut	CO <sub>2</sub> , H <sub>2</sub>	RF, NB, YE, CA, LI, TE	Leadbetter <i>et al.</i> , 1996
<i>Methanobrevibacter filiformis</i> <b>RFM-3<sup>T</sup></b>	Termite gut	CO <sub>2</sub> , H <sub>2</sub>	DTT, YE	Leadbetter <i>et al.</i> , 1998
<i>Methanobrevibacter acididurans</i> <b>ATM<sup>T</sup></b>	Anaerobic digester	CO <sub>2</sub> , H <sub>2</sub>	Acetate, RF	Savant <i>et al.</i> , 2002

<sup>1</sup>Type strain<sup>T</sup>, other strains classified under genus *Methanobrevibacter* in the NCBI database. <sup>v</sup>, strain variable. CoM, coenzyme M. RF, rumen fluid. VFA, volatile fatty acid. YE, yeast extract. TP, Trypticase peptone. DTT, dithiothreitol. CA, casamino acids. AA, amino acids. NB, nutrient broth. LI, liver infusion. TE, termite extract. Strains with sequenced genomes are displayed in bold.



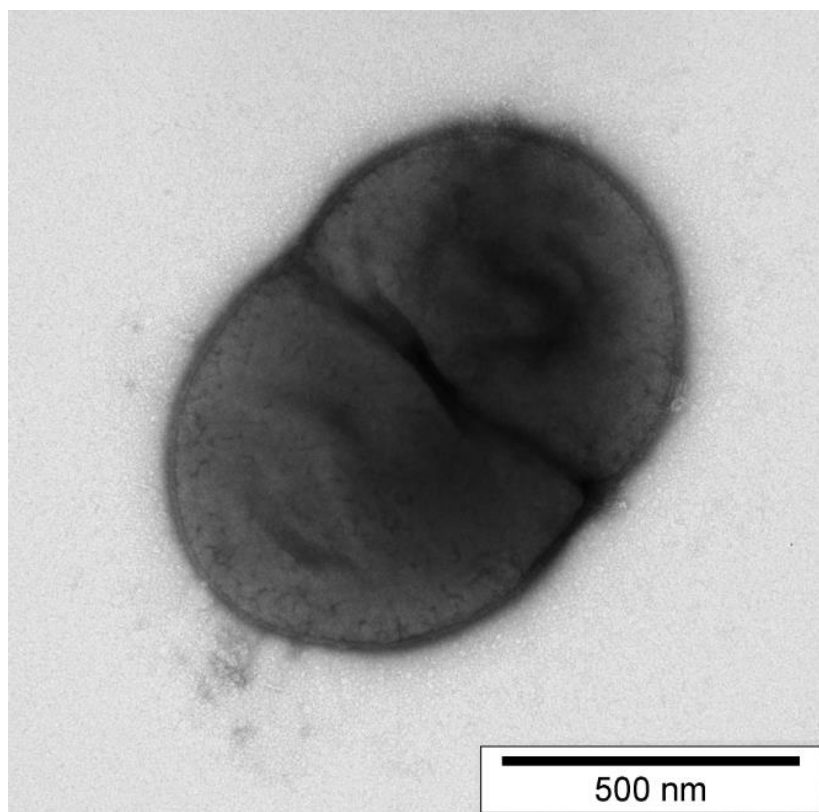
Members of the *Methanobrevibacter* genus identified from the rumen environment are classified phylogenetically according to 16S rRNA gene sequences into two major clades, the *Mbb. gottschalkii* clade and the *Mbb. ruminantium* clade. These clades account for 47% and 27% of rumen archaea, respectively, as reported in a global rumen microbial community survey based on 16S rRNA gene sequences (Henderson *et al.* 2015). In New Zealand rumen microbial populations, these two clades account for 42.4% and 32.9% of the rumen methanogen population (Seedorf *et al.* 2015). Thus far, 12 *Methanobrevibacter* species type strains and 27 other strains have been genome sequenced, the majority of these (21), reside within the *Mbb. smithii* clade (Table 6.1). There are 18 sequenced genomes included in this study, consisting of 11 type strains, five rumen strains, and two additional non-rumen strains were included from the *Mbb. arboriphilus* and *Mbb. oralis* clades.

In order to expand our understanding of members of the rumen *Mbb. gottschalkii* clade, an isolate from this clade was obtained. A methane-forming enrichment culture designated H6 was obtained from the rumen of a sheep (Tag number #J472) grazing a pasture diet (G. Henderson, personal communication). From this H6 culture a member of the *Mbb. gottschalkii* clade was isolated and designated as *Methanobrevibacter* sp. D5 (Kim 2012). In this chapter the genome sequence of *Methanobrevibacter* sp. D5 was assembled, annotated, analysed and compared to other available *Methanobrevibacter* genomes.

## 6.2 Results

### 6.2.1 Cell morphology

*Methanobrevibacter* sp. D5 has a short rod morphology which are 0.5 to 1.2  $\mu\text{M}$  in length and 0.4 to 0.7  $\mu\text{M}$  in width and is often observed as chains (Figure 6.1).



**Figure 6.1** Transmission electron micrograph of negatively stained thin section of the *Methanobrevibacter* sp. D5. The sample was prepared as previously described (Khan *et al.* 2013). Images were generated using a Philips CM10 Transmission Electron Microscope, and captured using an Olympus SIS Morada camera and SIS iTEM software (Germany).

### 6.2.2 Genome sequencing results and assembly

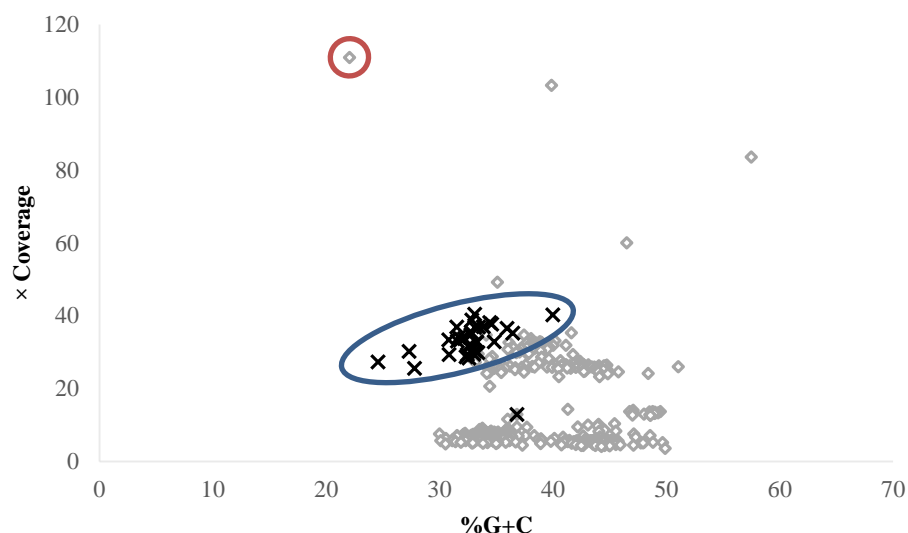
*Genome assembly.* Two *Methanobrevibacter* sp. D5 genome assemblies were performed during the course of this thesis; the first was performed using 454 pyrosequencing data from DNA isolated from the original H6 enrichment culture, and the second was performed using Illumina Miseq data, generated from DNA isolated from a pure culture of D5.

The sequencing of enrichment culture H6 yielded 1,407,413 reads with a total of 331,108,524 bp. The summary of the assembly statistics is displayed in Table 6.2. The assembly generated 383 scaffolds totaling 17,291,655 bp. Scaffolds belonging to the D5 genome were identified by sequence coverage, %G+C content and by the presence of ribosomal proteins. This analysis

resulted in the identification of 4 scaffolds (scaffolds 1, 15, 21 and 53) containing 40 contigs and totalling 2,756,038 bp (Figure 6.2). In addition to the D5 genome, DNA from five putative bacterial contaminants was identified. The bacterial contaminants were identified by 16S rRNA gene sequence identity as *Pseudobutyrvibrio ruminis* strain 153 (99% identity), *Olsenella umbonata* strain lac31 (99% identity), *Mogibacterium pumilum* strain D2-18 (93% identity), *Lachnospiraceae* bacterium strain DJF (94% identity) and *Erysipelotrichaceae* bacterium strain NK3D112 (93% identity).

**Table 6.2 Genome assembly summary of *Methanobrevibacter* sp. D5 from enrichment culture H6**

Genome sequence feature	Size or %G+C
<b>Enrichment culture H6</b>	
%G+C content	40.47%
Total size	15,210,579 bp
Coverage	18.6×
Number of reads	1,407,413
Number of bases	331,108,524 bp
Number of contigs	6,150
Number of scaffolds	383
Number of contigs >500 bp	4,882
N50 contig size	50,105
<b>Scaffold 1</b>	
%G+C content	32.94%
Size	2,270,142 bp
Number of reads	302,839
Number of contigs	32
<b>Scaffold 15</b>	
%G+C content	32.15%
Size	271,401 bp
Number of reads	39,317
Number of contigs	5
<b>Scaffold 21</b>	
%G+C content	33.85%
Size	160,014 bp
Number of reads	24,052
Number of contigs	2
<b>Scaffold 53</b>	
%G+C content	36.82%
Size	54,481 bp
Number of reads	2,817
Number of contigs	1
<b><i>Methanobrevibacter</i> sp. D5 total from the enrichment culture</b>	
%G+C content	33.1%
Size	2,756,038 bp
Number of contigs	40
Coverage	31.5×

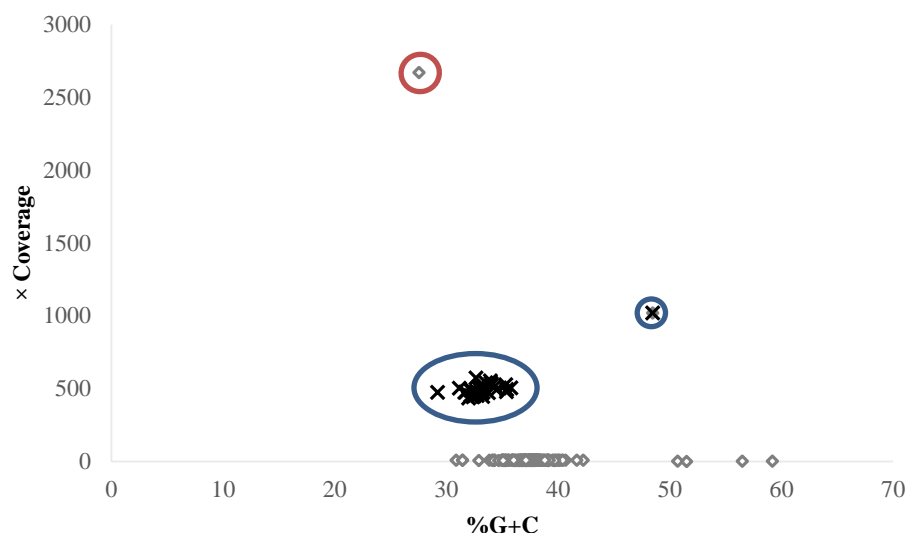


**Figure 6.2 Contig coverage and %G+C of enrichment culture H6 sequencing assembly.** (×) D5 contigs. (◇) Contigs of other organisms within the H6 enrichment culture. Only contigs above 3 kb are displayed. A putative prophage contig is circled in red. The cluster of D5 contigs are circled in blue.

The sequencing and assembly result of the *Methanobrevibacter* sp. D5 pure culture is summarised in Table 6.3. A hybrid assembly was produced by mapping the newly sequenced contigs onto the existing D5 assembly from the 454 assembly of the H6 enrichment culture. The Miseq assembly results indicated the presence of a second genome that had a low coverage in comparison to the *Methanobrevibacter* sp. D5 contigs (Figure 6.3). Based on the predicted ORFs and its BLASTP results, the contaminating genome likely belongs to a facultative anaerobe from the genus *Bacillus*. A D5 contig containing an rRNA operon has approximately double the coverage of other D5 contigs (Figure 6.3).

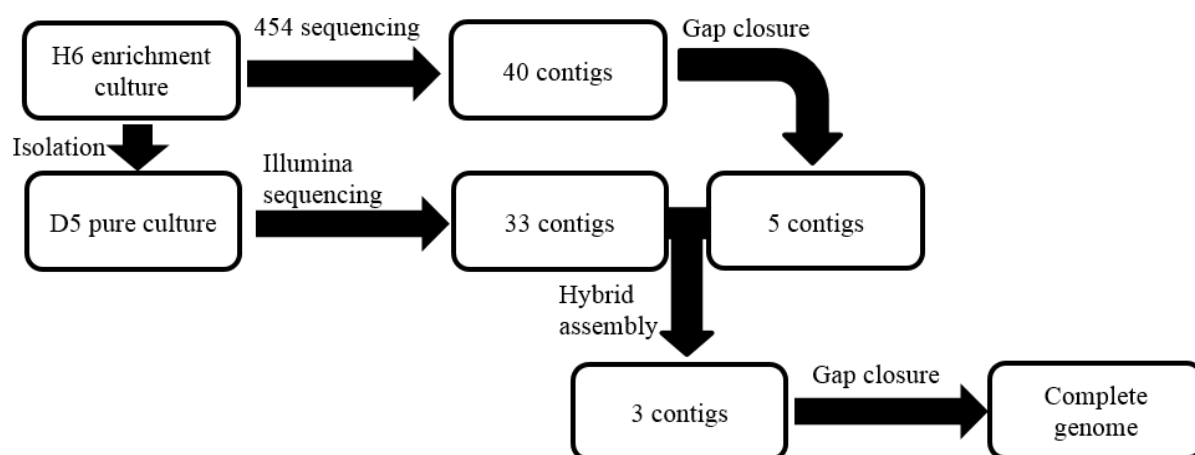
**Table 6.3 Genome assembly summary of *Methanobrevibacter* sp. D5**

Features	D5	Contaminating genome
Total size	2,826,590 bp	3,741,504 bp
Coverage	567.5	8.7
Number of total reads	13,667,636	
Number of total contigs	8770	
Number of contigs >6000 bp	104	
Number of D5 contigs	33	
N50 contig size	64,460 bp	39,481 bp
G+C content	33.1	37.4



**Figure 6.3 Contig coverage and %G+C of D5 pure culture sequencing assembly.** (x) D5 contigs. (◇) Contigs of other organisms. Only contigs above 3 kb are displayed. A putative prophage contig is circled in red. The cluster of D5 contigs are circled in blue.

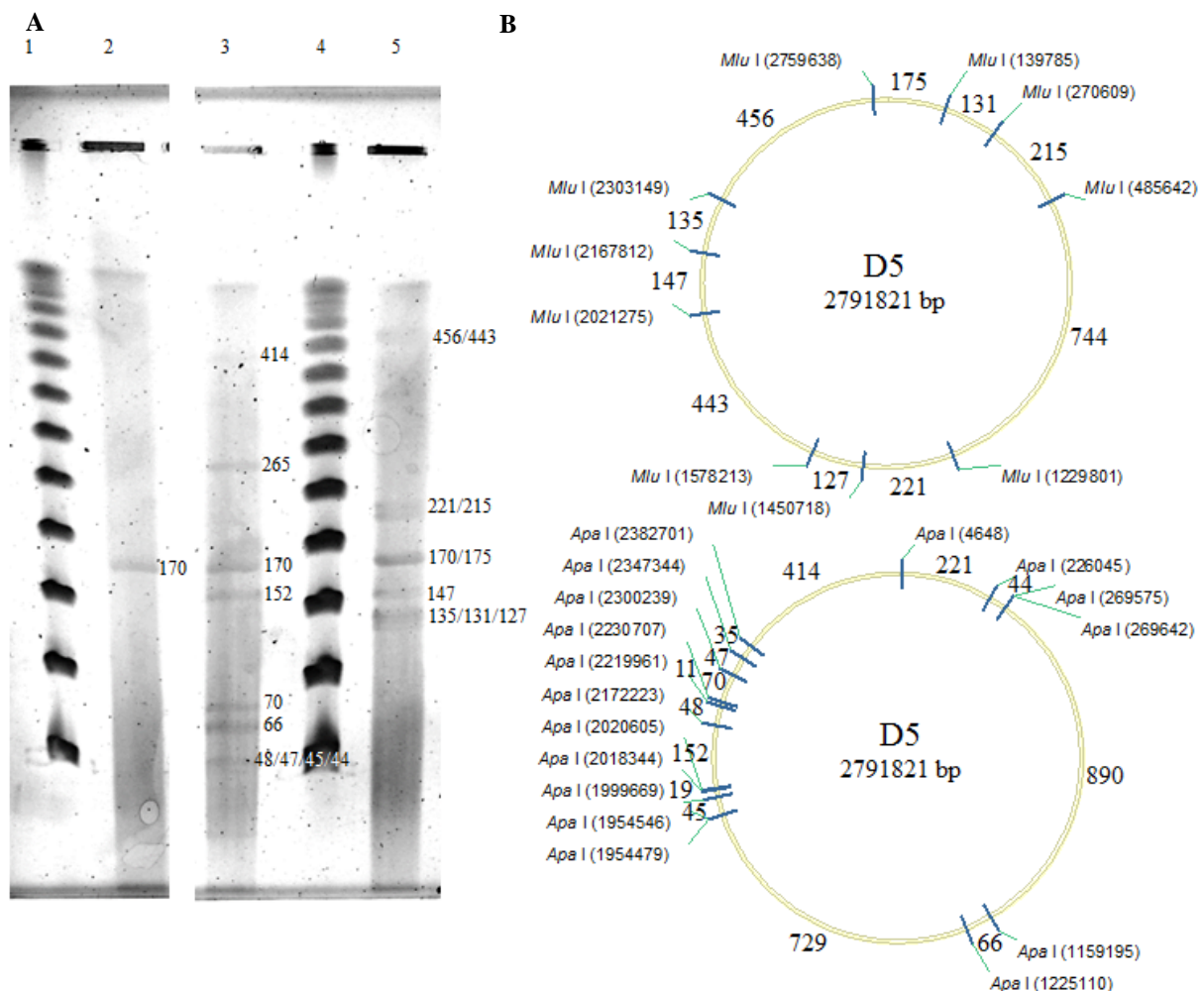
*Methanobrevibacter* sp. D5 Gap closure. To complete the *Methanobrevibacter* sp. D5 genome as a circular chromosome, it was necessary to close the sequence gaps between contigs. The outline of the D5 genome gap closure procedure is illustrated in Figure 6.4.



**Figure 6.4 Flow chart of D5 gap closure procedure.** Gap closure was performed on H6 assembly, and subsequently merged with Miseq assembly to create a hybrid assembly, further gap closure was performed on the hybrid assembly to close the genome.

The gap closure procedure is also described in Section 2.2.12. A total of 238 sequencing reactions were required to circularize the *Methanobrevibacter* sp. D5 genome as a single contig. Base conflicts in the hybrid assembly genome were validated by PCR amplification and re-sequencing.

*Methanobrevibacter* sp. D5 genome assembly verification. Assembly validation was confirmed by pulsed-field gel electrophoresis (Section 2.2.13). The restriction endonuclease *Mlu*I and *Apa*I were used (Figure 6.5). The *Mlu*I digestion pattern confirmed the assembly, whereas the *Apa*I digestion pattern did not. In the *Apa*I PFGE results, a 265 kb band appeared to be due to the absence of an *Apa*I restriction site at base 226,045, which caused the fused fragment of 221 kb and 44 kb. In order to assess this discrepancy, this region was validated by PCR amplification and re-sequenced and the presence of *Apa*I restriction endonuclease recognition site at base 226,045 was confirmed. Use of PFGE also revealed a 170 kb band which was observed in the digested and undigested lanes of both enzyme treatments in addition to the D5 genomic DNA.



**Figure 6.5** Pulse field gel electrophoresis of *Methanobrevibacter* sp. D5 genomic DNA. **A.** *Methanobrevibacter* sp. D5 genomic DNA digested with restriction endonucleases. Numbers beside each band indicates size in kb. Lane 1: concatenated lambda marker (fragment sizes, from bottom in kb, are 48.5, 97, 145.5, 194, 242.5, 291, 339.5, 388 and 436.5). Lane 2: undigested genomic DNA; Lane 3: digested with *Apa*I; Lane 4: concatenated lambda marker; Lane 5: digested with *Mlu*I. **B.** In silico restriction enzyme maps of *Mlu*I and *Apa*I

restriction endonuclease cleavage sites within the D5 genome. Numbers between restriction sites indicates fragment size in kb. Numbers in brackets indicates the base number of restriction sites.

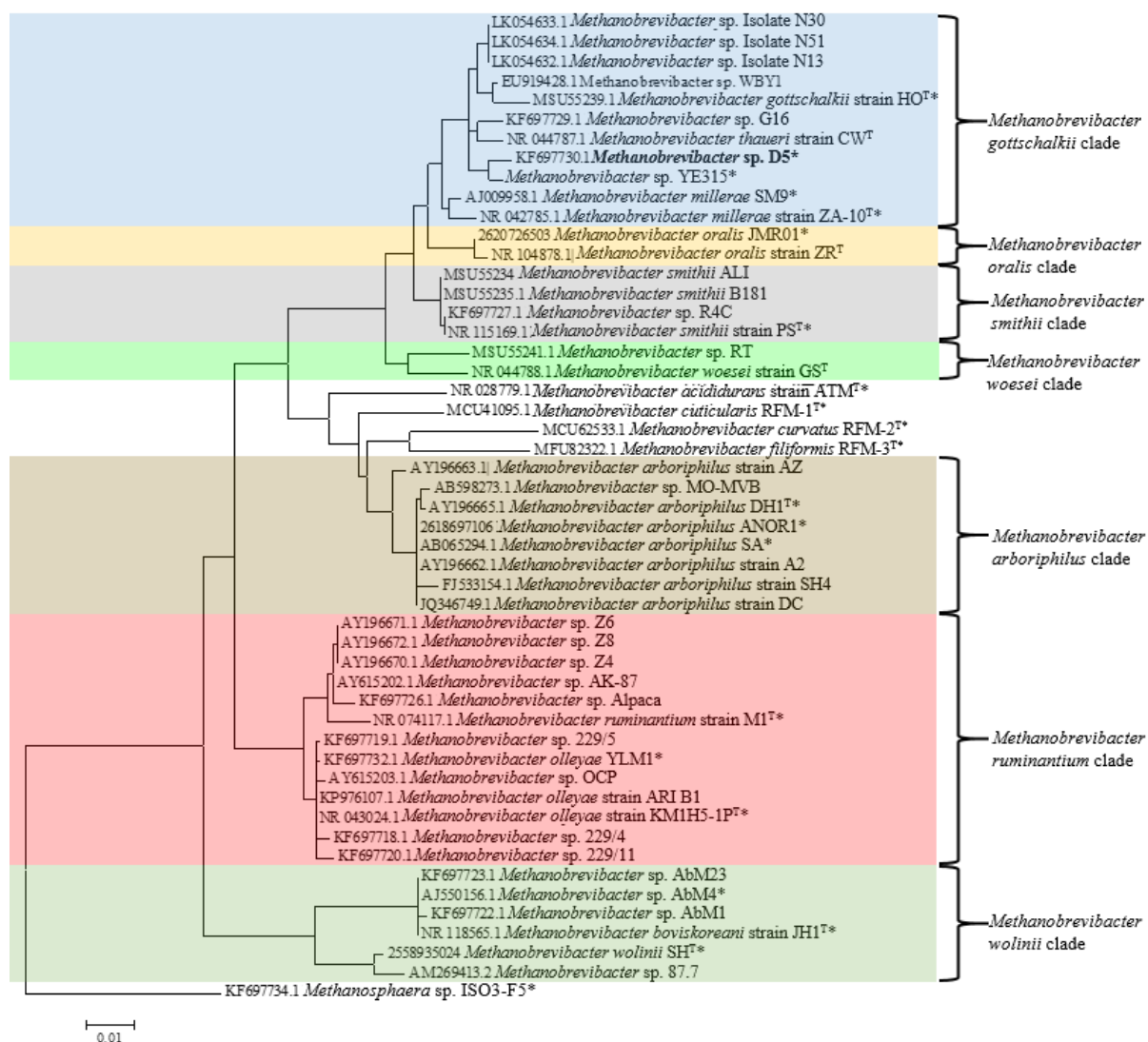
### 6.2.3 Phylogenetic relationship of sequenced members of *Methanobrevibacter*.

The phylogenic relationships of the members of the genus *Methanobrevibacter* was inferred from their 16S rRNA gene sequences (Table 6.4, Figure 6.6). A FGD tree (Figure 6.7) and a pan-genome tree (Figure 6.8) were created to assess the similarity between sequenced genomes based on the full complement of ORFs within each genome.

**Table 6.4 Phylogenetic divergence indicated by the % identity of 16S rRNA gene.**

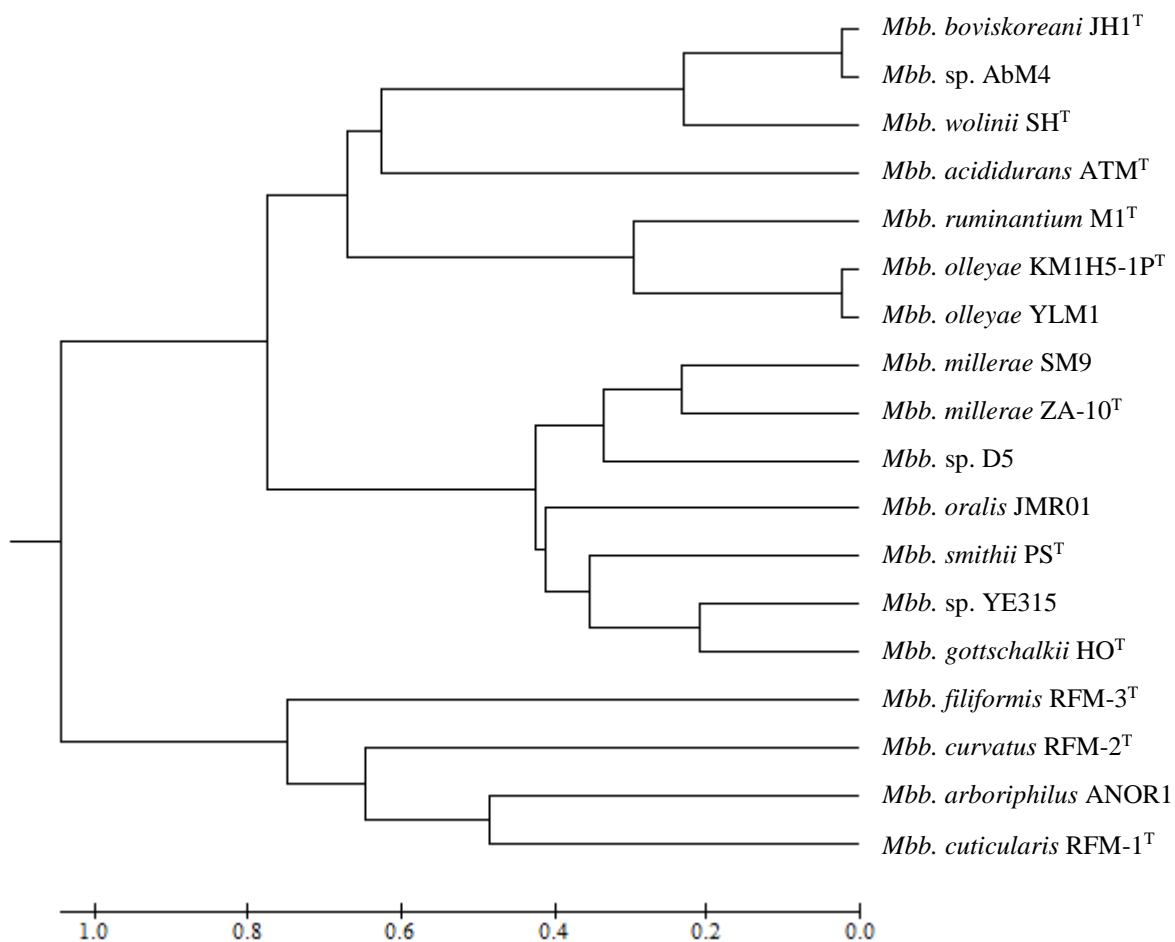
	AT M <sup>T</sup>	AN OR 1	JH1 T	RF M- 2 <sup>T</sup>	RF M- 1 <sup>T</sup>	RF M- 3 <sup>T</sup>	HO T	SM 9	Za- 10 <sup>T</sup>	KM IH5 - IP <sup>T</sup>	JM R01	M1 T	PS <sup>T</sup>	Ab M4	D5	YE 315	YL M1	SH <sup>T</sup>
AT M <sup>T</sup>		94.9	92.7	93.9	94.1	93.7	92	92.7	93	93.8	93.5	93.9	94	92.8	92.9	92.6	93.9	92.4
AN OR 1	94.9		93.2	94.8	96.4	94.7	92.6	93.6	94.2	94.4	93.7	93.1	94	92.5	93.4	93.6	94.4	92.6
JH1 T	92.7	93.2		91.6	91.7	92	91.7	92.3	92.3	93.2	92.4	93.2	93.4	<b>99.2</b>	92.6	92.6	93.4	96.3
RF M- 2 <sup>T</sup>	93.9	94.8	91.6		94.9	94.1	90.8	92.1	92.4	92.9	91.7	92.9	92.3	91	91.8	91.9	92.9	91.3
RF M- 1 <sup>T</sup>	94.1	96.4	91.7	94.9		94.3	91.8	93.3	93.2	93.4	93.1	93.5	93.4	91.7	92.9	93.1	93.6	91.6
RF M- 3 <sup>T</sup>	93.7	94.7	92	94.1	94.3		90.9	92.2	92.2	92.5	91.9	93.1	92.3	91.9	92.3	92.4	92.7	91.6
HO <sup>T</sup>	92	92.6	91.7	90.8	91.8	90.9		97.4	97.2	92.8	96.7	92.5	97	91.5	97.9	97.9	92.9	92.3
SM 9	92.7	93.6	92.3	92.1	93.3	92.2	97.4		<b>99.1</b>	93.2	97.4	92.8	98	91.9	<b>98.7</b>	98.5	93.5	92.2
Za- 10 <sup>T</sup>	93	94.2	92.3	92.4	93.2	92.2	97.2	<b>99.1</b>		93.2	97.2	93.1	97.9	92.4	<b>98.8</b>	<b>98.7</b>	93.4	92.5
KM IH5 -IP <sup>T</sup>	93.8	94.4	93.2	92.9	93.4	92.5	92.8	93.2	93.2		93.7	98.1	93.7	93	93.3	93.1	<b>99.9</b>	93.7
JM R01	93.5	93.7	92.4	91.7	93.1	91.9	96.7	97.4	97.2	93.7		92.9	98.2	92.2	97.4	96.9	93.6	92.2
M1 <sup>T</sup>	93.9	93.1	93.2	92.9	93.5	93.1	92.5	92.8	93.1	98.1	92.9		93.3	92.6	93	93.1	97.7	93.4
PS <sup>T</sup>	94	94	93.4	92.3	93.4	92.3	97	98	97.9	93.7	98.2	93.3		93.2	98	97.4	93.7	93.5
Ab M4	92.8	92.5	<b>99.2</b>	91	91.7	91.9	91.5	91.9	92.4	93	92.2	92.6	93.2		92.2	92.2	92.8	95.6
D5	92.9	93.4	92.6	91.8	92.9	92.3	97.9	<b>98.7</b>	<b>98.8</b>	93.3	97.4	93	98	92.2		<b>99</b>	93.6	92.8
YE3 15	92.6	93.6	92.6	91.9	93.1	92.4	97.9	98.5	<b>98.7</b>	93.1	96.9	93.1	97.4	92.2	<b>99</b>		93.5	92.5
YL M1	93.9	94.4	93.4	92.9	93.6	92.7	92.9	93.5	93.4	<b>99.9</b>	93.6	97.7	93.7	92.8	93.6	93.5		93.8
SH <sup>T</sup>	92.4	92.6	96.3	91.3	91.6	91.6	92.3	92.2	92.5	93.7	92.2	93.4	93.5	95.6	92.8	92.5	93.8	

Identities  $\geq 98.7\%$  are displayed in bold. All positions containing gaps or missing data were eliminated. Type strains are indicated by superscript T.

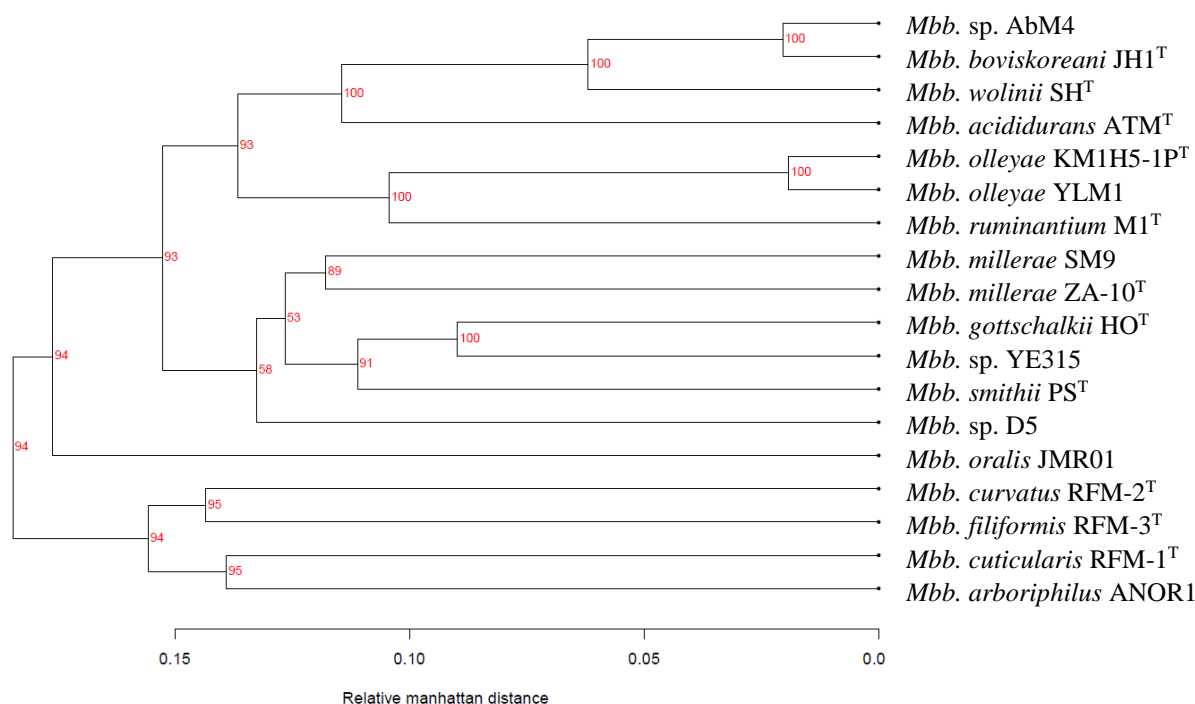


**Figure 6.6 16S rRNA gene-based phylogenetic tree.** The phylogenetic tree showing the relationships of cultured *Methanobrevibacter* spp. in relation to *Methanobrevibacter* sp. D5 (shown in bold print). The phylogeny was inferred from 16S rRNA gene nucleotide sequences (1,474 bp internal region) aligned using the Maximum Likelihood method based on the Kimura 2-parameter model. Bar: 0.01 substitutions per nucleotide position. The GenBank/JGI accession numbers of sequences are displayed. Strains whose genomes have been sequenced are marked with asterisk. The initial tree for the heuristic search was obtained by applying the Neighbour-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. All positions containing gaps or missing data were eliminated, giving a total of 667 positions in the final dataset. The 16S rRNA gene sequence from *Methanosphaera* sp. ISO3-F5 was used as an outgroup. Type strains are indicated by superscript T.





**Figure 6.7 Functional Genome Distribution tree based on of the *Methanobrevibacter* spp. ORFeome.** The evolutionary history was inferred using the UPGMA method (Sneath and Sokal 1962) based on the functional genome distribution (FGD) matrix (Altermann 2012). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetics tree. Type strains are indicated by superscript T.



**Figure 6.8 Pan genome tree of sequenced type strains and rumen members of *Methanobrevibacter* spp.** Pan genome family tree based on the absence and presence of gene families. The relative Manhattan distance indicates the proportion of the pan-genome where genomes differ in presence/absence of gene families. Bootstrap values from 1000 resamplings are indicated as percentages at each branch point. Type strains are indicated by superscript T.

Based on this analysis, *Methanobrevibacter* sp. D5 is most closely related to *Methanobrevibacter* sp. YE315, and most distantly related to *Mbb. arboriphilus* ANOR1 and the three type strains isolated from the termite hindgut. In this study, strains displaying a 16S rRNA gene identity of less than 98.7% are considered different species. Based on the 16S rRNA gene sequence identity, *Methanobrevibacter* sp. D5 can be considered as a strain of the *Mbb. millerae* species (Table 6.4). *Methanobrevibacter* sp. AbM4 is considered the same species as *Mbb. boviskoreani* JH1<sup>T</sup>. *Mbb. olleyae* YLM1 is considered the same species as *Mbb. olleyae* KM1H5-1P<sup>T</sup> (Table 6.4). The *Mbb. cuticularis* RFM-1<sup>T</sup>, *Mbb. curvatus* RFM-2<sup>T</sup>, *Mbb. filiformis* RFM-3<sup>T</sup> and *Mbb. acididurans* ATMT are affiliated to their own clades (Figure 6.6). Based on the FGD and pan genome tree, the *Mbb. smithii* PS<sup>T</sup> is similar to the members of the *Mbb. gottschalkii* clade (Figure 6.7, 6.8).

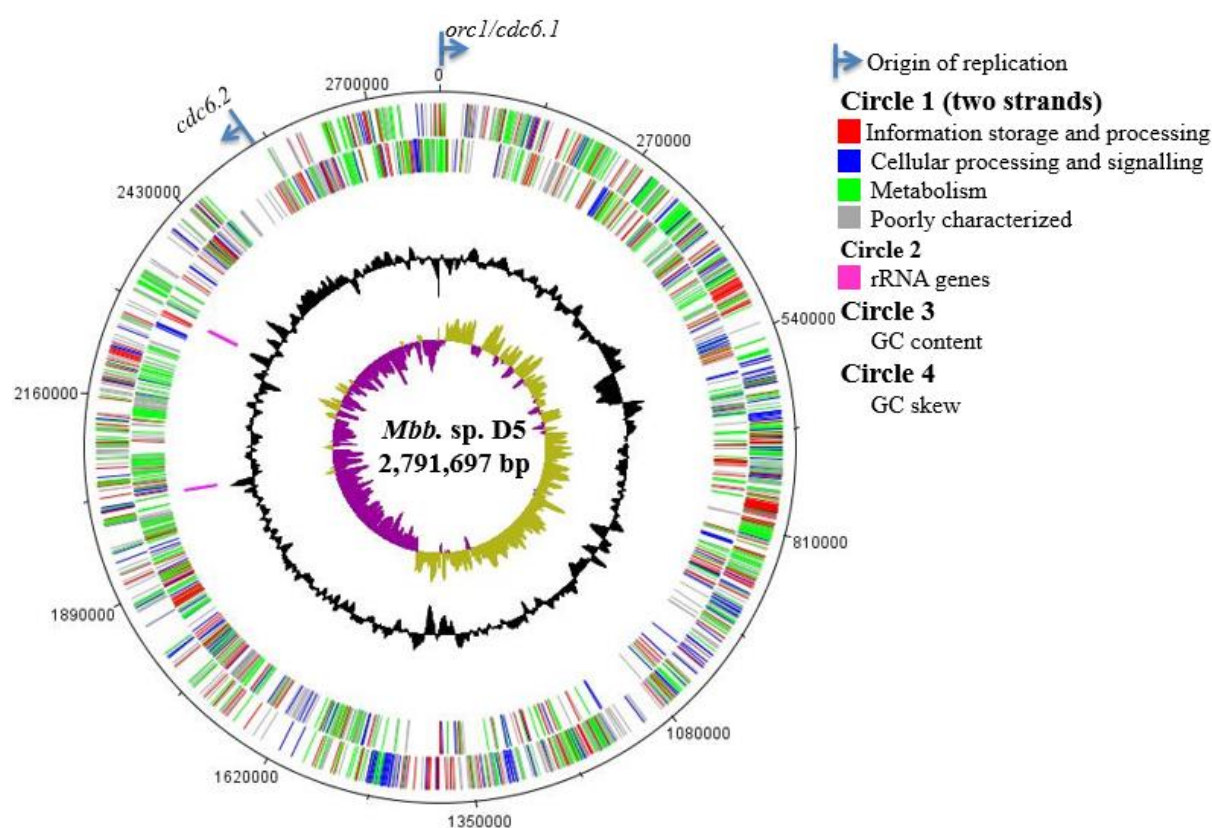
#### 6.2.4 Genome properties of *Methanobrevibacter* sp. D5 and other sequenced *Methanobrevibacter* spp.

The general features of the circular *Methanobrevibacter* sp. D5 genome are summarised in Table 6.5 and Figure 6.9. The D5 genome is 2,791,821 bp, 33.1 %G+C and encodes 2,487 ORFs. The genome was annotated and the functional characterisation of genes are summarised

in Figure 6.10. There are seven pseudogenes predicted in the genome of D5, including a pyridoxamine 5'-phosphate oxidase family protein (D5\_0093), a NADH oxidase (D5\_0115), an acetyltransferase GNAT family (D5\_0931), an ion transport protein (D5\_1201), a transporter  $\text{Na}^+/\text{H}^+$  antiporter (D5\_1397), a 5,10-methenyltetrahydromethanopterin hydrogenase Hmd (D5\_1879) and a potassium channel protein (D5\_2313).

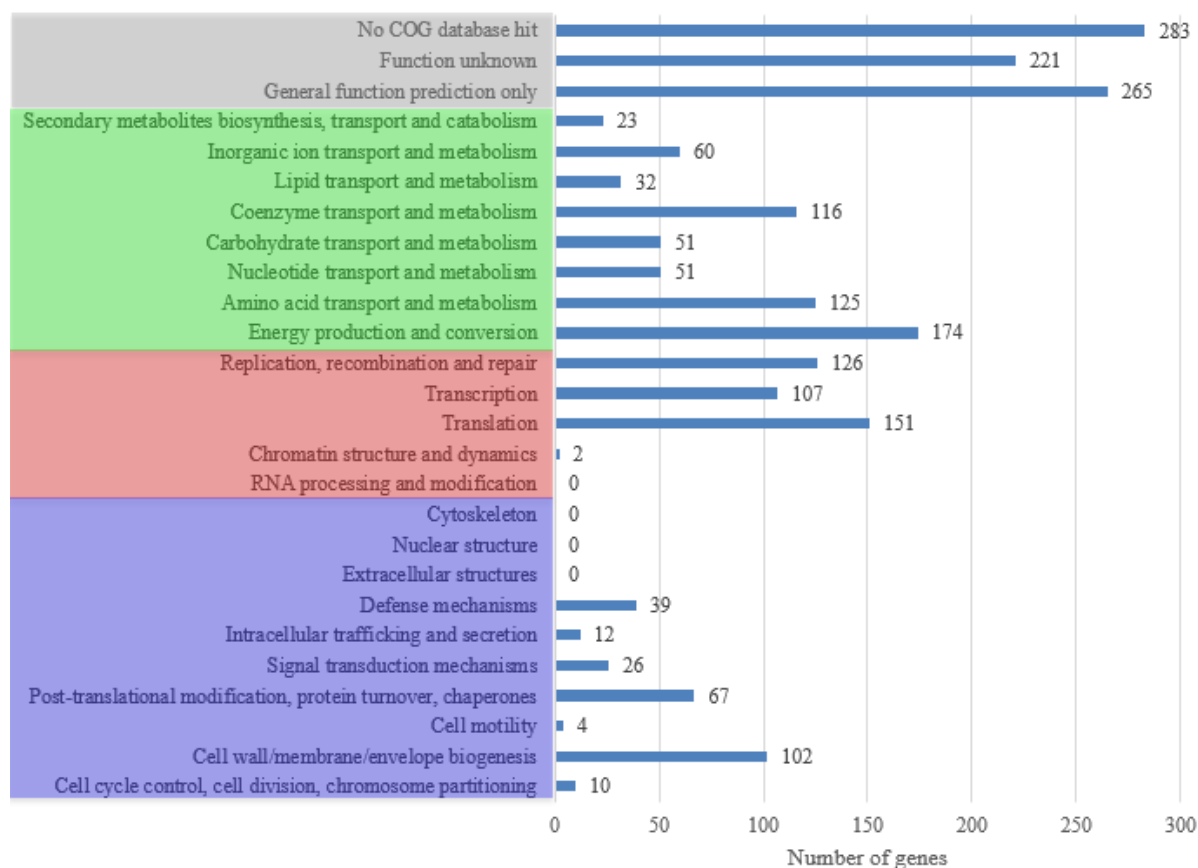
**Table 6.5 General features of the Methanobrevibacter sp. D5 genome**

Feature	
Source	Ovine rumen
Project status	Complete
Genome size (bp)	2,791,821
G+C content (%)	33.1
Number of ORFs	2,487
Coding area (%)	88.8
Contigs	1
rRNAs (5s, 16s, 23s)	3, 2, 2
tRNAs (with introns)	44 (2)
Non-coding RNAs	4
Insertion sequences	32
Prophage	No
CRISPR regions	1
Adhesin-like proteins	82



**Figure 6.9 Circular representation of the Methanobrevibacter sp. D5 genome.** Circles are referred to as 1 (outermost) to 4 (innermost). Circle 1: predicted ORFs on the + and – strands respectively. ORFs are coloured

based on the Clusters of Orthologous (COG) categories. Circle 2: location of the rRNA genes. Circle 3: %G+C content. Circle 4: GC bias  $[(G-C)/(G+C)]$ , khaki indicates values  $>1$ , purple  $<1$ . Origins of replication are represented as blue lines perpendicular to the outermost circle, with an arrow in direction of 5' to 3'.



**Figure 6.10 The functional classification *Methanobrevibacter* sp. D5.** The function classification of genes were predicted based on the clusters of orthologous proteins (COG) database (Tatusov *et al.* 2001).

*Genome characteristics of available Methanobrevibacter genomes.* The general genome characteristics of D5 in comparison to other sequenced *Mbb.* spp are listed in Table 6.6.

*Origin of replication.* The replication origin (*orc1/cdc6*) of the D5 chromosome was identified by GC nucleotide skew  $[(G-C)/(G+C)]$  analysis (Figure 6.9). Two *orc1/cdc6* genes were predicted in the genome (D5\_0001, D5\_2279). One Origin Recognition Box (ORB) motif was identified 87 bp upstream of the D5\_0001 gene. Therefore, the D5\_0001 gene is predicted to be the true origin of replication of D5.

In addition to D5, 12 other *Methanobrevibacter* genomes also host two copies of *cdc6* genes, whereas *Mbb. ruminantium* M1<sup>T</sup>, *Mbb. arboriphilus* ANOR1, *Methanobrevibacter* sp. AbM4, *Mbb. curvatus* RFM-2<sup>T</sup> and *Mbb. acididurans* ATM<sup>T</sup>, carry an additional truncated *cdc6* gene.

**Table 6.6 General genome properties of sequenced species of genus *Methanobrevibacter***

Species name	<i>Methanobrevibacter</i> sp. D5	<i>Methanobrevibacter</i> sp. YE315	<i>Methanobrevibacter</i> <i>milleriae</i> ZA-10 <sup>T</sup>	<i>Methanobrevibacter</i> <i>milleriae</i> SM9	<i>Methanobrevibacter</i> <i>gottschalkii</i> HO <sup>T</sup>	<i>Methanobrevibacter</i> <i>smithii</i> PS <sup>T</sup>
Source	Ovine rumen	Bovine rumen	Bovine rumen	Ovine rumen	Horse faeces, ovine rumen <sup>1</sup>	Sewage digester, human gut <sup>2</sup>
Project status	Complete	Complete	Draft	Complete	Draft	Complete
Genome size (bp)	2,791,697	2,273,296	2,725,667*	2,543,538	1,864,477*	1,853,160
G+C content (%)	33.1	34.2	36.4	31.8	30.2	31
ORFs	2,487	1,942	2,383	2,269	1,895	1,795
Coding area (%)	88.8	83.3	89.6	89	89.3	90.8
Contigs	1	1	48	1	20	1
rRNA operons	2	2	1	2	2	2
tRNAs (with introns)	44 (2)	33 (2)	77 (3)	40 (2)	35 (2)	34 (1)
Non-coding RNA	4	4	3	8	5	2
Insertion sequences	41	Nd	Nd	65	Nd	8
Prophage	No	Nd	Nd	Yes	Nd	Yes
CRISPR regions (spacers)	1 (20)	1 (6)	0*	2 (17, 20)	0*	1 (43)
Adhesin-like proteins	82	Nd	Nd	95	Nd	48
LPxTG-like motif	0	0	7	3	0	4
Sortase	1	1	1	1	1	1
NCBI accession		CP010834.1		NZ_CP011266.1		NC_009515.1
Publication	This study			Kelly <i>et al.</i> , 2016		Samuel <i>et al.</i> , 2007
Species name	<i>Methanobrevibacter</i> <i>ruminantium</i> M1 <sup>T</sup>	<i>Methanobrevibacter</i> <i>olleyae</i> KM1H5-1P <sup>T</sup>	<i>Methanobrevibacter</i> <i>olleyae</i> . YLM1	<i>Methanobrevibacter</i> <i>wolinii</i> SH <sup>T</sup>	<i>Methanobrevibacter</i> sp. AbM4	<i>Methanobrevibacter</i> <i>boviskoreani</i> JH1 <sup>T</sup>
Source	Bovine rumen	Ovine rumen	Ovine rumen	Sheep faeces, caprine rumen <sup>4</sup>	Ovine abomasum	Bovine rumen
Project status	Complete	Draft	Complete	Draft	Complete	Draft
Genome size (bp)	2,937,203	2,122,444*	2,201,192	2,041,814*	1,998,189	2,045,801
G+C content (%)	33	26.8	26.9	24.2	29	27.9
ORFs	2,115	1,813	1,839	1,700	1,671	1,756
Coding area (%)	81	76.9	75.8	76.3	75.8	78.4
Contigs	1	49	1	26	1	54
rRNA operons	2	1	2	2	3	1
tRNAs (with introns)	58 (2)	33 (2)	33 (1)	36 (2)	38 (2)	39 (2)
Non-coding RNA	3	4	5	3	4	3
Insertion sequences	4	Nd	19	Nd	23	1
Prophage	Yes	Nd	Yes	Nd	No	Nd
CRISPR regions (spacers)	3 (61, 11, 38)	1* (16)	2 (211, 21)	1* (4)	1 (246)	2* (139, 9)
Adhesin-like proteins	105	Nd	53	Nd	29	Nd
LPxTG-like motif	1	1	1	2	3	3
Sortase	1	1	1	1	1	1
NCBI accession	CP001719.1		CP014265.1	JHWX00000000.1	CP004050.1	BAGX00000000.2
Publication	Leahy <i>et al.</i> , 2010		Kelly <i>et al.</i> , 2016		Leahy <i>et al.</i> , 2013	Lee <i>et al.</i> , 2013

**Table 6.6 General genome properties of sequenced species of *Methanobrevibacter* genus (continued)**

Species name	<i>Methanobrevibacter cuticularis</i> RFM-1 <sup>T</sup>	<i>Methanobrevibacter curvatus</i> RFM-2 <sup>T</sup>	<i>Methanobrevibacter filiformis</i> RFM-3 <sup>T</sup>	<i>Methanobrevibacter acididurans</i> ATM <sup>T</sup>	<i>Methanobrevibacter oralis</i> JMR01	<i>Methanobrevibacter arboriphilus</i> ANOR1
Source	Termite gut	Termite gut	Termite gut	Anaerobic digester	Human faeces, human subgingival plaque <sup>3</sup>	Human faeces
Project status	Draft	Draft	Draft	Draft	Draft	Draft
Genome size (bp)	2,705,405*	2,600,395*	2,989,372*	1,655,232*	2,083,511*	2,221,072*
G+C content (%)	26.9	28.8	27.03	27.3	27.8	25.5
ORFs	2,154	2,149	2,517	1,575	2,251	1,993
Coding area (%)	71.6	76.3	76.2	85.2	82.4	74
Contigs	190	192	460	36	60	5
rRNA operons	1	1	1	1	1	2
tRNAs (with introns)	35 (2)	34 (2)	33 (2)	33 (2)	33 (2)	35 (2)
Non-coding RNA	2	2	3	3	3	3
Insertion sequences	Nd	Nd	Nd	Nd	Nd	Nd
Prophage	Nd	Nd	Nd	Nd	No	No
CRISPR regions (spacers)	3 (13,6,72)	1 (50)	3* (61, 5, 3)	1* (163)	3* (33, 5, 21)	3 (2, 2, 5)*
Adhesin-like proteins	Nd	Nd	Nd	Nd	Nd	Nd
LPxTG-like motif	0	0	0	1	1	0
Sortase	1	1	1	1	1	1
NCBI accession	LWMW000000000.1	LWMV000000000.1	LWMT000000000.1		CBWS000000000.1	NZ_CBVX000000000.1
Publication					Khelaifia <i>et al.</i> , 2014	Khelaifia <i>et al.</i> , 2014

\*Among currently available sequences. In addition to the environment which the sequenced organism was initially obtained, the organism has also been identified in other environments, as indicated by 1. (Jeyanathan *et al.* 2011). 2. (Matarazzo *et al.* 2012). 3. (Irbis and Ushida 2004). 4. (Ferrari *et al.* 1994). #incomplete regions

*Ribosomal, non-coding and transfer RNAs.* The ribosomal RNAs in completely sequenced members of the genus *Methanobrevibacter* have been found in the characteristic operon arrangement, 16S, 23S and 5S rRNA genes. An additional copy of the 5S rRNA gene is present in *Metanobrevibacter* sp. D5, *Mbb. ruminantium* M1<sup>T</sup>, *Mbb. boviskoreani* JH1<sup>T</sup>, *Mbb. smithii* PS<sup>T</sup>, *Mbb. millerae* ZA-10<sup>T</sup>, *Mbb. olleyae* KM1H5-1P<sup>T</sup>, *Metanobrevibacter* sp. YE315, *Mbb. arboriphilus* ANOR1, *Mbb. curvatus* RFM-2<sup>T</sup>, *Mbb. filiformis* RFM-3<sup>T</sup> and *Mbb. acididurans* ATM<sup>T</sup>, while two additional copies of the 5S rRNA gene are predicted in the genomes of *Mbb. gottschalkii* HO<sup>T</sup>, *Mbb. oralis* JMR01 and *Mbb. olleyae* YLM1.

Four non-coding RNAs (ncRNAs) are predicted in the *Methanobrevibacter* sp. D5 genome, including an archaeal RNaseP, an archaeal signal recognition particle, a group II catalytic intron and a FMN riboswitch. The ncRNAs predicted within the other *Mbb.* genomes analysed are summarised in Table 6.7. All of the protein components of RNaseP and the signal recognition particle have been predicted in the genomes analysed.

Seventeen of the *Methanobrevibacter* genomes analysed are predicted to contain all of the tRNA genes for 20 amino acids (Table 6.8), but the tRNA<sup>Ala</sup> is not detected in the *Mbb. olleyae* KM1H5-1P<sup>T</sup> draft genome. The numbers of tRNAs corresponding to each amino acid is highly conserved between the genomes analysed, with the exception of *Mbb. millerae* ZA-10<sup>T</sup> genome and *Mbb. ruminantium* M1<sup>T</sup> genome, which contains a comparatively larger number of tRNAs. Introns have been found consistently within the tRNA genes tRNA<sup>Trp</sup> and tRNA<sup>Met</sup> across all the *Methanobrevibacter* genomes analysed, with an additional intron in the tRNA<sup>Ser</sup> gene of *Mbb. millerae* genome and the tRNA<sup>Lys</sup> gene of *Mbb. ruminantium* M1<sup>T</sup> genome.

**Table 6.7 Non-coding RNAs predicted to be present in *Methanobrevibacter* genomes**

Features	D5	YE315	ZA-10 <sup>T</sup>	SM9	HO <sup>T</sup>	PS <sup>T</sup>	JMR01	M1 <sup>T</sup>	KM1H5-1P <sup>T</sup>	YLM1	SH <sup>T</sup>	AbM4	JH1 <sup>T</sup>	RFM-1 <sup>T</sup>	RFM-2 <sup>T</sup>	RFM-3 <sup>T</sup>	ANOR1	ATM <sup>T</sup>
Archaeal RNaseP	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Archaeal Signal recognition particle	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Group II catalytic intron	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
FMN riboswitch (RFN element)	+	+	+	+		+	+	+	+	+	+	+	+	-	-	-	-	+
HNH endonuclease associated RNA	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
crcB RNA	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-
microRNA mir-598	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-

+ indicates the predicted presence of the ncRNA, - indicates the predicted absence of the ncRNA

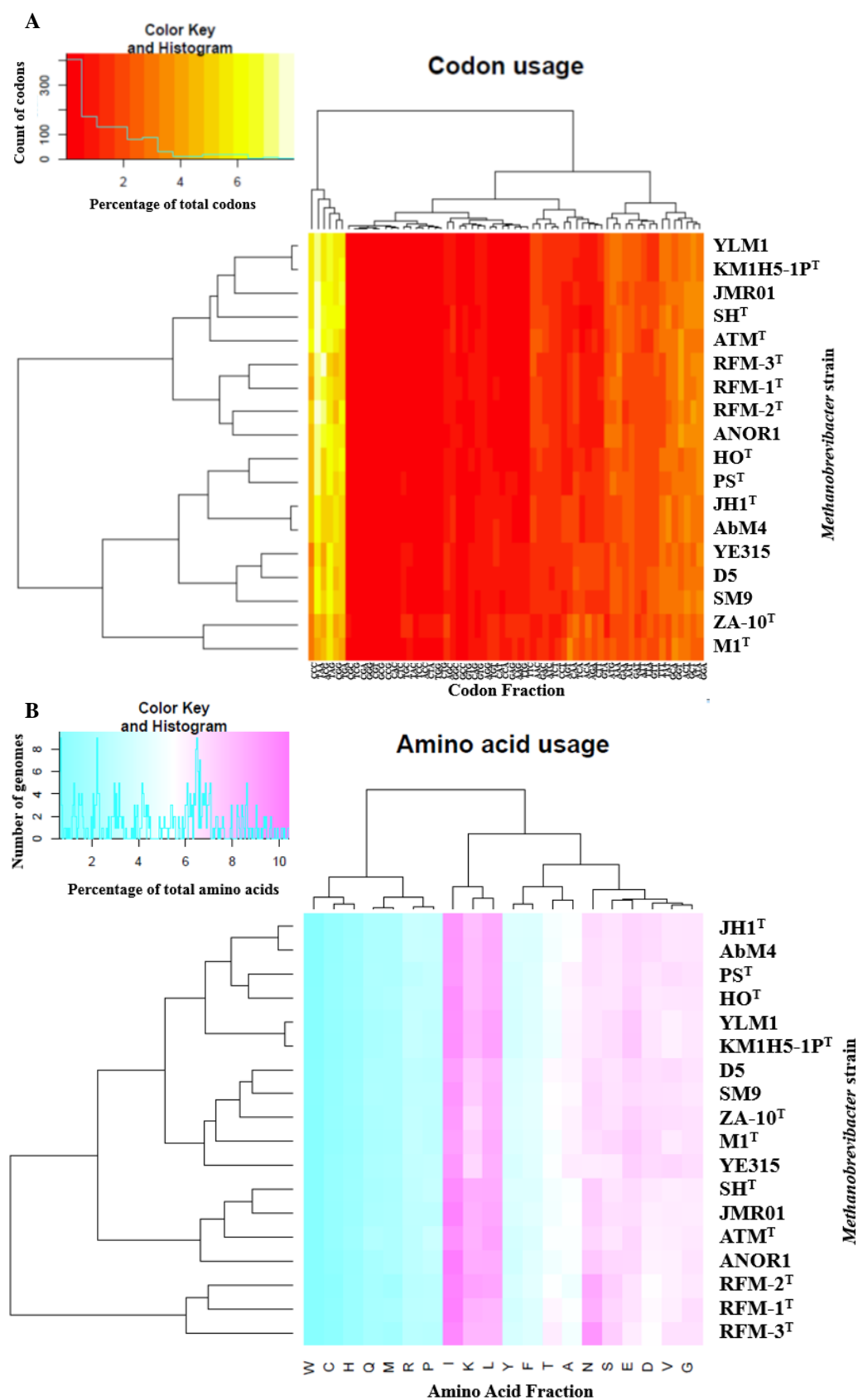
**Table 6.8 Predicted tRNAs in *Methanobrevibacter* spp. genomes**

tRNA corresponding Amino acid	D5	YE315	ZA-10 <sup>T</sup>	SM9	HO <sup>T</sup>	PS <sup>T</sup>	JMR01	M1 <sup>T</sup>	KM1H5-1P <sup>T</sup>	YLM1	SH <sup>T</sup>	AbM4	JH1 <sup>T</sup>	RFM-1 <sup>T</sup>	RFM-2 <sup>T</sup>	RFM-3 <sup>T</sup>	ANOR1	ATM <sup>T</sup>
Alanine	2	2	2	2	2	2	2	4	0	2	2	2	2	2	2	2	3	2
Arginine	4	4	6	4	4	4	4	5	4	4	3	4	4	5	4	4	4	3
Asparagine	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Aspartate	1	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
Cysteine	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Glutamine	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1
Glutamate	1	1	3	1	2	1	1	2	1	1	3	2	2	1	1	1	1	1
Glycine	2	2	4	2	2	2	2	3	2	2	2	2	2	2	2	2	2	2
Histidine	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Isoleucine	2	1	2	1	1	1	1	2	1	1	1	2	2	1	1	1	1	1
Leucine	5	3	8	4	3	3	3	6	3	3	3	3	3	3	3	3	3	3
Lysine	2	1	4	2	1	1	1	4*	1	1	1	2	2	1	1	1	1	2
Methionine	3*	3*	7*	3*	3*	3*	3*	3*	3*	3*	3*	3*	3*	4**	3*	3*	3*	3*
Phenylalanine	2	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	2	1
Proline	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
Serine	4	3	6*	4	3	5	3	6	3	3	3	3	3	3	3	3	3	3
Threonine	2	2	6	4	2	3	2	2	2	2	2	2	2	2	2	2	2	2
Tryptophan	1*	1*	2*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*
Tyrosine	2	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
Valine	3	2	4	2	2	2	2	4	3	3	2	2	2	2	2	2	2	2
Pseudo	2	0	11	2	1	0	0	6	2	0	2	2	2	0	0	0	0	0
Total	44	33	77	40	35	34	33	58	33	33	36	38	39	35	34	33	35	33
Intron containing tRNAs	2	2	3	2	2	2	2	3	2	2	2	2	2	2	2	2	2	2

\*intron-containing tRNAs



Codon and amino acid usage of *Methanobrevibacter* species. The codon usage for each genome is summarised in Table A.6.2 and visualised in Figure 6.11.



**Figure 6.11 Codon and amino acid usage of the *Methanobrevibacter* genomes analysed.** **A.** Codon usage heatmap of the *Methanobrevibacter* genomes analysed. The color key on the top left hand corner, indicates codons with the lowest percentage usage in red, and most abundant codons are displayed in pale yellow. The histogram displays the count of codons in 0.37% intervals across genomes. **B.** Amino acid usage heatmap of the *Methanobrevibacter* genomes analysed. The heatmap was clustered to place organisms with the shortest distance under the same branch. The color key displayed in this heatmap is illustrated on the top left hand corner, the most scarce amino acid is displayed in cyan, and most abundant amino acid is displayed in magenta. The number of genomes was tallied at intervals of 0.05% and displayed in a histogram. The figure was generated by CMG-Biotools (Vesth *et al.* 2013).

The *Methanobrevibacter* genomes display a high degree of homogeneity in codon and amino acid usage. The codon usage of ZA-10<sup>T</sup> and M1<sup>T</sup> are slightly different in comparison to other *Methanobrevibacter* analysed, as reflected in the codon usage for the amino acids leucine and isoleucine (Table A.6.2). The RFM-1<sup>T</sup>, RFM-2<sup>T</sup>, RFM-3<sup>T</sup> genomes are predicted to use a higher proportion of asparagine. Translational start codons are similar among the genomes investigated (Table A.6.2). The KM1H5-1P<sup>T</sup>, YLM1, SH<sup>T</sup>, AbM4, JH1<sup>T</sup> and ANOR1 genomes have a stronger preference for using the opal codon as a stop codon (68.2% to 74.1%) than the remainder of the genomes analysed (57.5% to 64.9%) (Table A.6.2).

*Horizontal gene transfer of Methanobrevibacter sp. D5.* Two different softwares were used to predict horizontal gene transfer, Alien Hunter (Vernikos and Parkhill, 2006) was used to identify horizontally transferred regions independent of genome annotation. There are 80 regions predicted by Alien Hunter to contain horizontally acquired DNA based on atypical sequence composition (Table 6.9), and these regions account for 30.7% of the D5 genome. A total of 535 genes are predicted across the 80 regions. Region 30 had a moderate likelihood of being horizontally transferred.

**Table 6.9 Predicted horizontal gene transfer regions of *Methanobrevibacter* sp. D5**

Region	Base range	IVOM score*	%G+C	Number of genes	Average CAI*	Likelihood
1	12500..17500	24.818	35.2	1	0.779	Low
2	52500..82500	20.738	34.2	7	0.679	Low
3	150000..165000	29.713	36.0	3	0.694	Low
4	175000..180000	17.788	32.4	6	0.690	Low
5	225000..230000	17.331	36.4	6	0.802	Low
6	247500..252500	16.185	32.4	2	0.794	Low
7	282500..292500	19.14	36.3	4	0.677	Low
8	350000..360000	37.63	37.4	13	0.793	Low
9	365000..370000	22.739	34.2	4	0.809	Low
10	465000..492500	26.629	37.3	24	0.801	Low
11	495000..500000	19.895	36.3	7	0.589	Low
12	502500..512500	25.64	36.7	8	0.606	Low
13	515000..525000	27.473	37.9	4	0.639	Low
14	535000..540000	19.643	24.0	3	0.793	Low
15	547500..570000	21.184	26.2	13	0.768	Low
16	572500..590000	25.181	25.7	8	0.768	Low
17	597500..627500	29.052	37.8	28	0.622	Low
18	630000..635000	18.735	37.4	3	0.606	Low
19	655000..662500	17.717	35.8	5	0.656	Low
20	710000..715000	20.107	26.1	2	0.776	Low
21	772500..787500	20.949	33.0	26	0.817	Low
22	797500..802500	18.459	31.8	9	0.839	Low
23	815000..820000	15.563	33.7	4	0.835	Low

24	887500..892500	23.403	38.2	4	0.635	Low
25	895000..900000	24.79	38.6	4	0.655	Low
26	910000..917500	18.885	25.8	5	0.799	Low
27	935000..942500	19.447	37.5	7	0.677	Low
28	945000..950000	21.946	39.1	4	0.695	Low
29	952500..957500	17.069	35.5	3	0.637	Low
30	972500..987500	46.045	35.3	7	0.652	Moderate
31	992500..997500	18.446	34.8	2	0.675	Low
32	1000000..1007500	18.759	30.1	6	0.725	Low
33	1042500..1047500	15.802	33.6	3	0.638	Low
34	1062500..1070000	18.391	37.7	5	0.699	Low
35	1080000..1112500	21.618	34.2	3	0.762	Low
36	1115000..1135000	26.752	35.0	1	0.765	Low
37	1187500..1197500	17.395	35.0	3	0.714	Low
38	1225000..1232500	17.675	35.5	8	0.652	Low
39	1252500..1260000	26.053	34.7	2	0.750	Low
40	1275000..1280000	16.692	34.8	5	0.661	Low
41	1330000..1342500	17.087	35.6	16	0.635	Low
42	1345000..1362500	24.993	37.1	21	0.603	Low
43	1380000..1397500	18.324	27.2	12	0.754	Low
44	1402500..1410000	21.206	36.0	7	0.574	Low
45	1420000..1437500	29.047	26.8	11	0.688	Low
46	1440000..1462500	24.752	37.2	25	0.599	Low
47	1525000..1530000	16.604	36.2	3	0.668	Low
48	1580000..1592500	30.052	34.7	2	0.757	Low
49	1595000..1600000	16.993	35.3	2	0.730	Low
50	1627500..1632500	21.721	33.6	12	0.709	Low
51	1655000..1662500	18.642	33.7	3	0.748	Low
52	1665000..1675000	22.437	33.6	2	0.781	Low
53	1800000..1805000	17.586	36.3	3	0.626	Low
54	1810000..1830000	39.015	34.2	1	0.785	Low
55	1862500..1870000	18.655	31.1	4	0.809	Low
56	1932500..1942500	16.788	31.4	6	0.737	Low
57	1992500..2000000	21.047	38.2	6	0.731	Low
58	2015000..2025000	60.635	39.3	4	0.736	Low
59	2140000..2145000	24.285	36.7	4	0.797	Low
60	2195000..2200000	16.104	36.6	3	0.710	Low
61	2245000..2252500	21.284	25.74	6	0.801	Low
62	2255000..2260000	22.598	35.9	3	0.681	Low
63	2262500..2270000	18.925	36.7	11	0.717	Low
64	2295000..2307500	47.987	39.1	6	0.689	Low
65	2340000..2350000	20.18	37.6	10	0.670	Low
66	2352500..2357500	17.623	33.9	2	0.766	Low
67	2370000..2385000	45.87	39.5	5	0.774	Low
68	2417500..2427500	16.264	36.7	9	0.726	Low
69	2445000..2452500	18.62	37.1	8	0.675	Low
70	2460000..2530000	32.628	36.8	5	0.738	Low
71	2532500..2542500	29.914	37.7	1	0.708	Low
72	2550000..2555000	16.571	37.5	8	0.697	Low
73	2560000..2567500	19.277	36.4	10	0.665	Low
74	2572500..2577500	22.203	36.0	2	0.618	Low
75	2597500..2602500	15.903	36.6	9	0.662	Low
76	2627500..2632500	18.747	37.4	2	0.649	Low
77	2647500..2665000	27.298	36.2	16	0.730	Low
78	2675000..2685000	22.538	37.4	5	0.699	Low
79	2730000..2737500	19.002	37.9	3	0.698	Low
80	2742500..2750000	17.388	33.1	5	0.737	Low

\*Atypical sequence composition is indicated by interpolated variable ordered motif (IVOM) score, average CAI (codon adaptation index) and %G+C. IVOM scores is annotation independent and it implemented variable order *k*-mers as reliable estimates of local sequence composition, as higher order motifs are more likely to capture deviation from genome background. A high regional IVOM score, abnormal % G+C and low average CAI is associated with a high likelihood of horizontal transfer.

Region 30 has a 35.3% G+C content close to the average 33.1% G+C content of the genome. This region has a high IVOM score with slightly low CAI. This region encodes six hypothetical proteins and a transposase.

*Regularly Interspaced Short Palindromic Repeat (CRISPR) element.* The *Methanobrevibacter* sp. D5 genome contains several CRISPR-associated genes (D5\_0810, 0880, 0881, 1219 and 1929) with a CRISPR region containing 19 repeats (bases 1,381,409 to 1,382,733). The direct repeat is 29 bp in length with the consensus sequence GTTTAAAATAGACTTAATAGTATGGAAAT. The presence of CRISPR-associated genes *cas1*, 2, 4, 6 and TIGR02710 suggested the D5 CRISPR/Cas system belongs to type U. A total of 20 predicted spacer sequences with species specific protospacer adjacent motif (PAM) were used to identify the potential targets of CRISPR RNAs (crRNAs). The predicted crRNA targets are displayed in Table 6.10. Spacers with no database matches (spacers 1, 6, 7, 8, 9, 15, 17, 18, 19) were omitted.

**Table 6.10 CRISPR spacer homology**

Spacer Number	Spacer sequence match*	Accession number	Score
2, 20	<i>Providencia rettgeri</i> strain 09ACRGNY2001 plasmid pPrY2001	NC_022589	22
3	Cyanophage P-RSM1	HQ634175	18
4	<i>Escherichia</i> phage JMPW2	KU194205	18
5	<i>Bacillus cereus</i> strain S2-8 plasmid pBFR_2	NZ_CP009606	18
10	Periwinkle leaf yellowing phytoplasma plasmid p09PLY-1	NC_019245	23
11	<i>Listeria</i> phage LP-048	KJ094033	20
12	<i>Zymomonas mobilis</i> subsp. Mobilis str. CP4 = NRRL B-14023 plasmid pZM1402301	NZ_CP003716	16
13	<i>Candidatus pantoea carbekii</i> strain US plasmid pBMSBPS3	NZ_CP010910	18
14	<i>Synechococcus</i> phage S-SSM5	GU071097	22
16	<i>Clostridium perfringens</i> strain JP838 plasmid pJFP838A	NZ_CP013615	17

\* Spacer sequence identified by BLAST based homology screening (Biswas *et al.* 2013). Score was calculated by matches (+1) and mismatches (-1) across the whole length of the spacer without gaps.

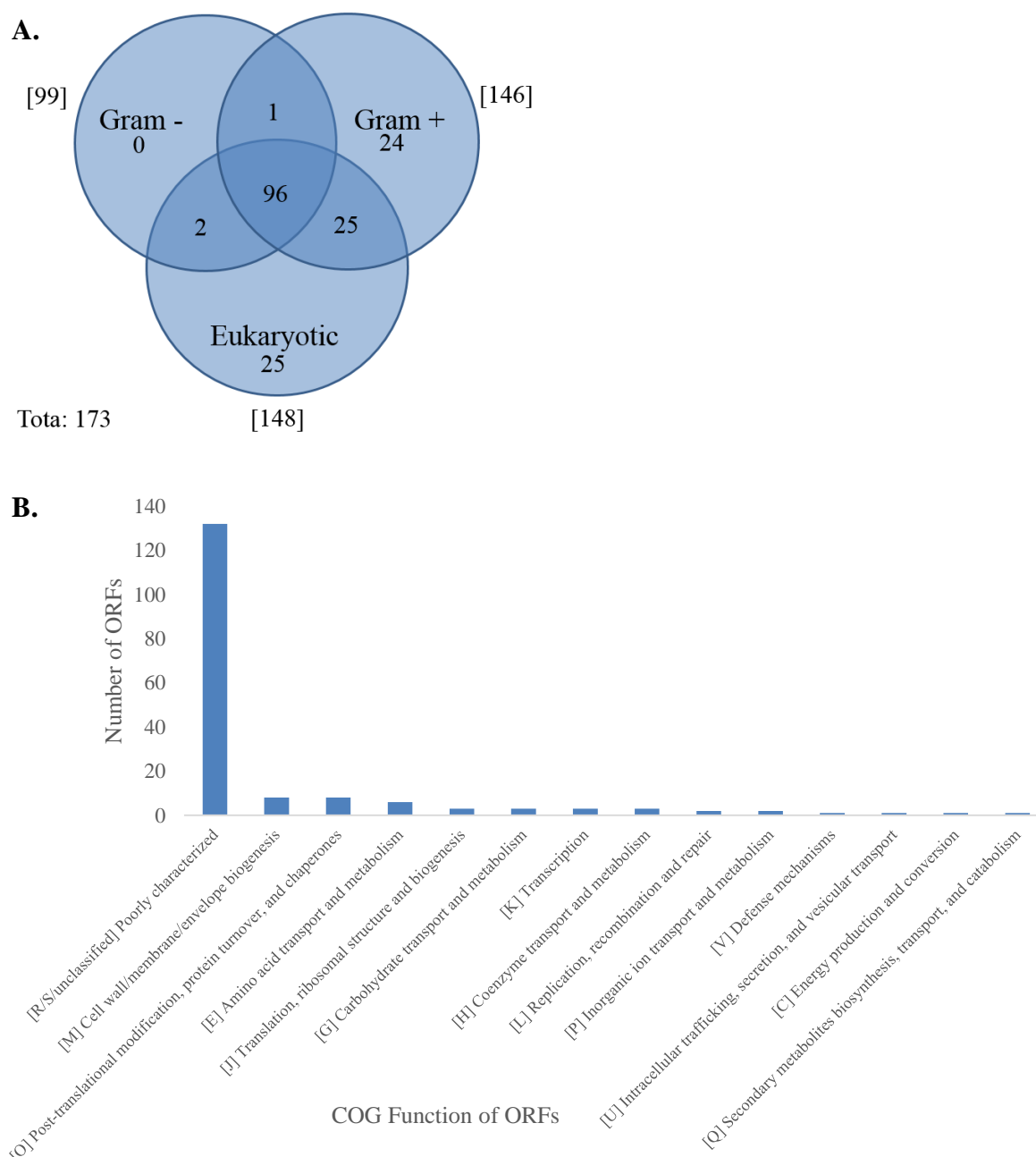
The CRISPR elements were predicted in 16 of the genomes analysed, containing from 1 to 3 CRISPR elements, with 4 to 246 spacer sequences (Table 6.6). No CRISPR elements have been predicted in the genomes of ZA-10<sup>T</sup> and HO<sup>T</sup>. The CRISPR associated genes have been predicted and are summarised in Table A.6.3.

*Insertion sequence elements (IS elements) of Methanobrevibacter sp. D5.* Insertion sequence elements were annotated as described in Section 2.2.22. A total of 32 IS elements were identified in the genome of *Methanobrevibacter sp. D5*, and these elements account for 1.8% of the coding DNA sequence. Two types of insertion sequences are found; transposons and phage integrases. The IS elements found in *Methanobrevibacter sp. D5* belong to families IS200/IS605, IS1380, IS1182, IS5, IS630 and ISNCY, with the IS200/IS605 family predicted with the highest copy number (Table 6.11).

**Table 6.11 Insertion sequence elements of *Methanobrevibacter sp. D5***

Locus tag	Family	Closest aligned source
D5_0467	IS1182	<i>Bacillus cereus</i>
D5_0936	IS1182	<i>Uncultured archaeon</i>
D5_1082	IS1182	<i>Methanosarcina acetivorans</i>
D5_1480	IS1182	<i>Microscilla sp.</i>
D5_0440	IS1380	<i>Rhodococcus opacus</i>
D5_0439	IS200/IS605	<i>Caldicellulosiruptor saccharolyticus</i>
D5_0441	IS200/IS605	<i>Methanosarcina acetivorans</i>
D5_0442	IS200/IS605	<i>Methanosarcina mazei</i>
D5_0506	IS200/IS605	<i>Caldicellulosiruptor saccharolyticus</i>
D5_0829	IS200/IS605	<i>Enterococcus faecium</i>
D5_0830	IS200/IS605	<i>Enterococcus faecium</i>
D5_1102	IS200/IS605	<i>Halobacillus halophilus</i>
D5_1177	IS200/IS605	<i>Halobacillus halophilus</i>
D5_1330	IS200/IS605	<i>Halobacillus halophilus</i>
D5_1331	IS200/IS605	<i>Enterococcus faecium</i>
D5_1622	IS200/IS605	<i>Halobacillus halophilus</i>
D5_1994	IS200/IS605	<i>Enterococcus faecium</i>
D5_2272	IS200/IS605	<i>Methanosarcina mazei</i>
D5_0892	IS5	<i>Methanobrevibacter ruminantium</i>
D5_2002	IS5	<i>Methanohalophilus mahii</i>
D5_2474	IS5	<i>Methanohalobium evestigatum</i>
D5_0019	IS630	<i>Archaeoglobus fulgidus</i>
D5_1402	IS630	<i>Archaeoglobus fulgidus</i>
D5_1403	IS630	<i>Uncultured archaeon</i>
D5_1766	IS630	<i>Archaeoglobus fulgidus</i>
D5_0495	ISNCY	<i>Methanobrevibacter smithii</i>
D5_0709	ISNCY	<i>Methanobrevibacter smithii</i>
D5_0710	ISNCY	<i>Methanobrevibacter smithii</i>
D5_0711	ISNCY	<i>Methanobrevibacter smithii</i>
D5_1223	ISNCY	<i>Methanobrevibacter smithii</i>
D5_1505	ISNCY	<i>Methanobrevibacter smithii</i>
D5_2341	ISNCY	<i>Methanobrevibacter smithii</i>

*Secretome*. The *Methanobrevibacter* sp. D5 genome encodes 173 ORFs predicted to contain a signal peptide (Figure 6.12A). Among the ORFeome proteins that could be exported, most ORF functions have been classified under the poorly characterized COG category (Figure 6.12B). Some archaea utilise ESCRT-III (eukaryotic endosomal complex required for transport) protein for vesicle secretion, but no homologues of ESCRT-III subunits involved in vesicle secretion were detected in the *Methanobrevibacter* sp. D5 genome.



**Figure 6.12 Distribution of *Methanobrevibacter* sp. D5 signal peptides.** A. Number of signal peptide ORFs classified according to SignalP training model. B. COG classification of D5 ORFs predicted with signal peptide.

The predicted secretome (surface associated and/or released from the cell) proteins of the eighteen *Methanobrevibacter* genomes are summarised in Table 6.12 and Table A.6.4. The genomes analysed are predicted to devote between 3.5% and 7% of their ORFs to the export of extracellular and surface-associated proteins. A large proportion of the secretomes examined are predicted to have ORFs with motifs or domains for attachment to the cell surface. Amongst the genomes analysed, between 64 to 147 genes per genome are predicted to be membrane anchored by one to two transmembrane helices or by an N-terminal lipobox. Around 3 to 27 genes per genome are predicted to be either unattached (i.e. secreted) or associated with the cell wall using other (unknown) mechanisms. The extracellular proteins were analysed for the presence of repeat domains and periplasmic domains (the periplasmic domain belongs to the substrate binding subunit of ABC transporters) and the results are listed in Table A.6.4.

**Table 6.12 Predicted secretome in *Methanobrevibacter* species**

Secretome	Integral membrane protein	Lipobox	LPxTG	Two TMH Membrane anchor	C-terminal membrane anchor	N-terminal membrane anchor	Secreted	Total (% of total ORFs)
D5	13	9	0	20	2	103	26	173 (7)
YE315	11	10	0	19	1	65	15	121 (6.2)
ZA-10 <sup>T</sup>	11	18	7	7	3	82	27	155 (6.5)
SM9	9	16	3	16	5	67	19	135 (5.9)
HO <sup>T</sup>	9	6	0	11	1	47	9	83 (4.4)
PS <sup>T</sup>	3	5	4	11	5	41	8	77 (4.2)
JMR01 <sup>T</sup>	7	4	1	7	0	68	13	100 (4.4)
M1 <sup>T</sup>	9	8	1	10	1	90	14	133 (6.3)
KM1H5-1P <sup>T</sup>	7	3	1	11	3	47	10	82 (4.5)
YLM1	9	4	1	15	4	46	11	90 (4.9)
SH <sup>T</sup>	6	4	2	18	1	26	3	60 (3.5)
AbM4	7	3	3	12	3	28	4	60 (3.6)
JH1 <sup>T</sup>	8	3	3	11	2	33	5	65 (3.7)
RFM-1 <sup>T</sup>	10	12	0	21	4	106	23	176 (8.2)
RFM-2 <sup>T</sup>	5	13	0	12	2	80	12	124 (5.8)
RFM-3 <sup>T</sup>	5	11	0	14	1	73	22	126 (5.0)
ATM <sup>T</sup>	7	9	0	3	1	38	7	65 (4.1)
ANOR1	6	1	0	13	1	43	12	76 (3.8)

Prediction E value cutoff  $1 \times 10^{-05}$ .

*Adhesin-like proteins of Methanobrevibacter sp. D5.* There are 81 genes predicted to be adhesin-like proteins in the D5 genome and these are displayed in Table 6.13, showing their predicted subcellular location, molecular weight, transmembrane helices, signal peptide and protein domains. Four adhesin-like proteins are predicted to contain a transglutaminase domain (D5\_0392, 0657, 1669, 1926), and one adhesin-like protein is predicted to contain a collagen-binding surface protein (D5\_1479).

**Table 6.13 Adhesin-like protein prediction**

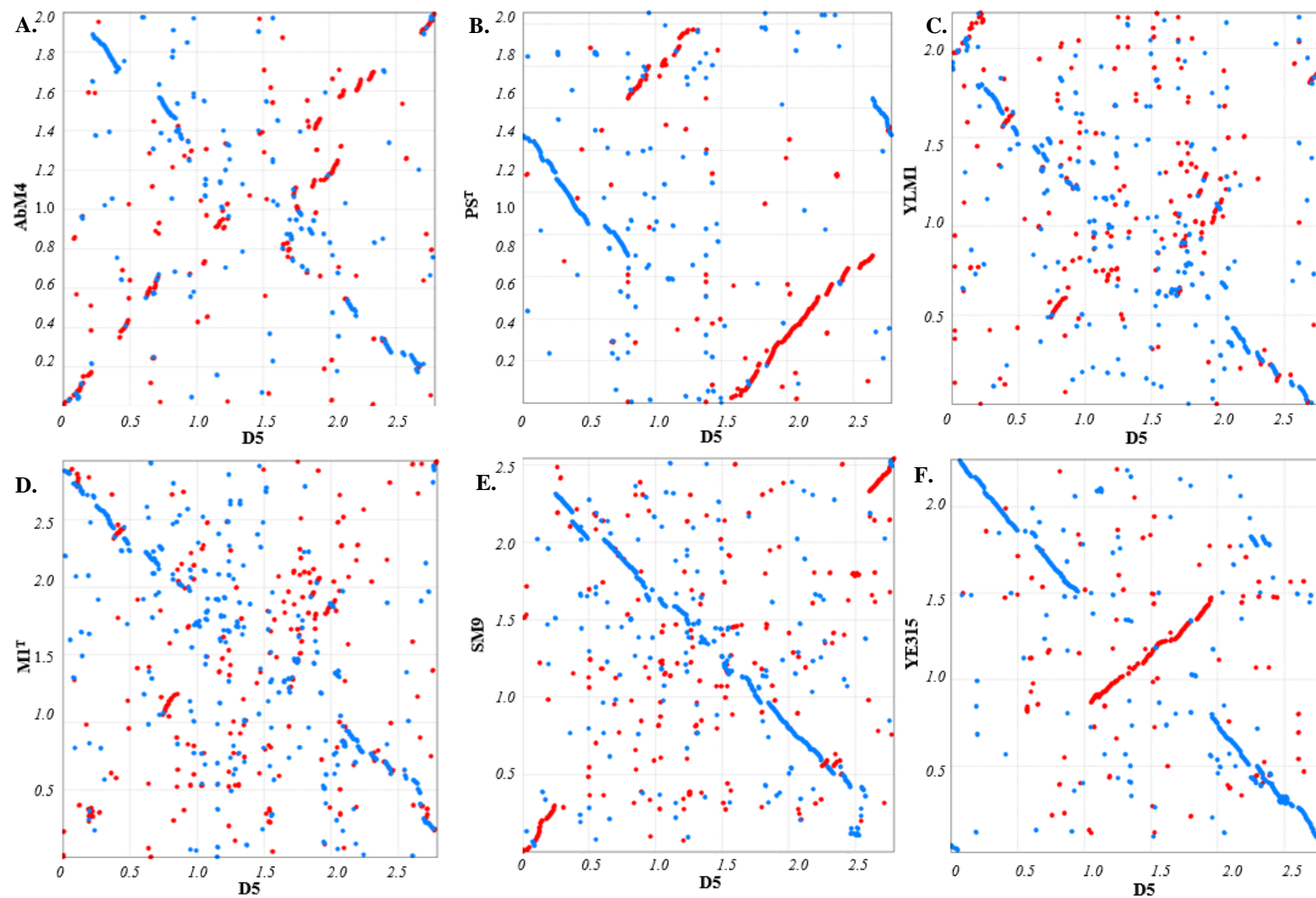
Locus_tag	Size (kDa <sup>1</sup> )	TMH <sup>2</sup>	SignalP	SPAAN P <sub>ad</sub> -value	Predicted gene product	LOCTree3	PSORTb	SubLoc	Domain
D5_0014	<b>210.13</b>	Yes	Yes	0.92	Adhesin-like protein	Secreted	Extracellular	Extracellular	Pectin lyase fold
D5_0015	<b>242.85</b>	Yes	Yes	0.88	Adhesin-like protein	Cytoplasm	Cellwall	Extracellular	Bacterial Ig-like, domain 1
D5_0016	<b>206.08</b>	Yes	No	0.90	Adhesin-like protein	Secreted	Extracellular	Periplasmic	Pectin lyase fold
D5_0036	<b>353.28</b>	Yes	Yes	0.92	Adhesin-like protein	Secreted	Cellwall	Periplasmic	Domain of unknown function DUF11
D5_0038	<b>466.91</b>	Yes	Yes	0.98	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_0051	22.80	Yes	Yes	0.94	Adhesin-like protein	Secreted	Extracellular	Extracellular	Bacterial Ig-like, domain 1
D5_0111	131.09	Yes	Yes	0.84	Adhesin-like protein	Cytoplasm	Cellwall	Extracellular	Pectin lyase fold
D5_0112	<b>555.26</b>	Yes	Yes	0.93	Adhesin-like protein	Secreted	Cellwall	Cytoplasmic	Pectin lyase fold
D5_0127	106.98	Yes	Yes	0.78	Adhesin-like protein	Secreted	Unknown	Extracellular	Bacterial Ig-like, domain 1
D5_0173	16.91	Yes	Yes	0.89	Adhesin-like protein	Secreted	Extracellular	Extracellular	
D5_0188	155.41	Yes	Yes	0.85	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_0189	136.01	No	No	0.77	Adhesin-like protein	Secreted	Unknown	Extracellular	Carbohydrate-binding-like fold
D5_0218	188.19	Yes	Yes	0.86	Adhesin-like protein	Secreted	Extracellular	Extracellular	Pectin lyase fold
D5_0219	78.79	Yes	Yes	0.77	Adhesin-like protein	Secreted	Cytoplasmic	Extracellular	Pectin lyase fold
D5_0252	20.19	Yes	Yes	0.83	Adhesin-like protein	Plasma membrane	Unknown	Extracellular	
D5_0262	81.64	Yes	Yes	0.72	Adhesin-like protein	Plasma membrane	Unknown	Extracellular	Bacterial Ig-like, domain 1
D5_0392	22.21	No	No	0.80	Adhesin-like protein	Secreted	Extracellular	Extracellular	Bacterial Ig-like, domain 1
D5_0393	28.28	No	No	0.84	Adhesin-like protein	Secreted	Cytoplasmic	Extracellular	Bacterial Ig-like, domain 1
D5_0394	30.02	No	No	0.80	Adhesin-like protein	Secreted	Extracellular	Extracellular	Transglutaminase-like
D5_0452	183.19	Yes	Yes	0.77	Adhesin-like protein	Secreted	Cellwall	Extracellular	CAP domain
D5_0453	93.69	Yes	Yes	0.70	Adhesin-like protein	Secreted	Extracellular	Extracellular	Bacterial Ig-like, domain 1
D5_0458	69.68	Yes	Yes	0.70	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_0459	106.59	Yes	Yes	0.79	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_0478	19.38	Yes	No	0.78	Adhesin-like protein	Secreted	Unknown	Cytoplasm	
D5_0640	33.91	Yes	Yes	0.79	Adhesin-like protein	Secreted	Cellwall	Extracellular	Bacterial Ig-like, domain 1
D5_0660	94.35	No	No	0.87	Adhesin-like protein	Secreted	Unknown	Extracellular	Transglutaminase-like, Bacterial Ig-like, domain 1
D5_0715	84.41	No	Yes	0.77	Adhesin-like protein	Secreted	Extracellular	Extracellular	Pectin lyase fold
D5_0821	58.59	No	No	0.86	Adhesin-like protein	Secreted	Extracellular	Extracellular	Pectin lyase fold
D5_0825	52.82	Yes	Yes	0.90	Adhesin-like protein	Secreted	Extracellular	Extracellular	P22 tailspike C-terminal domain-like
D5_0867	67.09	Yes	Yes	0.84	Adhesin-like protein	Secreted	Extracellular	Extracellular	Carboxypeptidase, regulatory domain
D5_0965	26.38	Yes	Yes	0.71	Adhesin-like protein	Secreted	Extracellular	Extracellular	
D5_0970	<b>387.18</b>	Yes	Yes	0.90	Adhesin-like protein	Secreted	Cellwall	Extracellular	Quinonprotein alcohol dehydrogenase-like superfamily
D5_0971	<b>674.93</b>	Yes	Yes	0.94	Adhesin-like protein	Plasma membrane	Cellwall	Extracellular	Pectin lyase fold
D5_0973	<b>257.92</b>	Yes	Yes	0.83	Adhesin-like protein	Plasma membrane	Cellwall	Extracellular	Peptidase C1A
D5_0974	<b>569.24</b>	Yes	Yes	0.97	Adhesin-like protein	Cytoplasm	Cellwall	Extracellular	Pectin lyase fold
D5_0976	188.57	Yes	Yes	0.86	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_0978	90.69	Yes	Yes	0.81	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_0979	84.77	Yes	Yes	0.72	Adhesin-like protein	Secreted	Unknown	Extracellular	
D5_1071	38.67	Yes	Yes	0.71	Adhesin-like protein	Secreted	Unknown	Extracellular	Bacterial Ig-like, domain 1
D5_1092	<b>251.97</b>	No	Yes	0.96	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_1152	93.71	Yes	Yes	0.70	Adhesin-like protein	Secreted	Unknown	Extracellular	Domain of unknown function DUF3344
D5_1322	42.03	Yes	Yes	0.72	Adhesin-like protein	Secreted	Unknown	Extracellular	



D5_1393	<b>214.15</b>	No	Yes	0.83	Adhesin-like protein	Secreted	Unknown	Extracellular	Peptidase C1A
D5_1400	19.34	Yes	Yes	0.93	Adhesin-like protein	Secreted	Extracellular	Extracellular	
D5_1406	<b>438.14</b>	No	Yes	0.88	Adhesin-like protein	Secreted	Cellwall	Extracellular	Filamin repeat
D5_1411	154.47	Yes	Yes	0.84	Adhesin-like protein	Secreted	Unknown	Extracellular	Peptidase C1A
D5_1412	94.65	No	Yes	0.84	Adhesin-like protein	Secreted	Cellwall	Extracellular	Bacterial Ig-like, domain 1
D5_1413	168.85	Yes	Yes	0.87	Adhesin-like protein	Secreted	Unknown	Extracellular	Peptidase C1A
D5_1423	126.25	Yes	Yes	0.84	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_1424	186.87	Yes	Yes	0.81	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_1473	<b>239.48</b>	Yes	Yes	0.77	Adhesin-like protein	Secreted	Cellwall	Extracellular	Collagen-binding surface protein
D5_1476	<b>440.17</b>	Yes	Yes	0.96	Adhesin-like protein	Secreted	Cellwall	Extracellular	Bacterial Ig-like, group 2
D5_1494	44.9	No	No	0.71	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_1610	30.23	Yes	Yes	0.84	Adhesin-like protein	Secreted	Extracellular	Periplasmic	Domain of unknown function DUF1002
D5_1618	134.96	Yes	Yes	0.86	Adhesin-like protein	Secreted	Unknown	Extracellular	Peptidase C1A
D5_1619	<b>687.06</b>	Yes	Yes	0.97	Adhesin-like protein	too long	Cellwall	Extracellular	Domain of unknown function DUF11
D5_1662	142.47	Yes	Yes	0.91	Adhesin-like protein	Secreted	Unknown	Extracellular	Transglutaminase-like, Bacterial Ig-like, domain 1
D5_1689	22.57	Yes	No	0.78	Adhesin-like protein	Secreted	Extracellular	Periplasmic	Zinc-ribbon domain
D5_1715	62.73	Yes	Yes	0.79	Adhesin-like protein	Cytoplasm	Unknown	Cytoplasmic	Pectin lyase fold
D5_1716	105.69	Yes	Yes	0.86	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_1746	13.68	Yes	No	0.81	Adhesin-like protein	Secreted	Extracellular	Extracellular	Immunoglobulin-like fold
D5_1758	72.15	Yes	No	0.71	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_1764	67.74	Yes	Yes	0.76	Adhesin-like protein	Secreted	Cytoplasmic	Extracellular	Peptidase C39
D5_1896	37.36	No	Yes	0.70	Adhesin-like protein	Secreted	Cellwall	Extracellular	
D5_1912	<b>206.63</b>	Yes	Yes	0.76	Adhesin-like protein	Secreted	Unknown	Periplasmic	Domain of unknown function DUF11
D5_1921	104.46	Yes	Yes	0.87	Adhesin-like protein	Secreted	Cytoplasmic	Extracellular	Transglutaminase-like, Bacterial Ig-like, domain 1
D5_2120	14.17	Yes	Yes	0.86	Adhesin-like protein	Secreted	Unknown	Extracellular	
D5_2160	38.47	No	No	0.70	Adhesin-like protein	Secreted	Unknown	Extracellular	Bacterial Ig-like, domain 1
D5_2173	194.14	No	Yes	0.78	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_2175	<b>201.09</b>	Yes	Yes	0.77	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_2176	<b>231.5</b>	Yes	Yes	0.75	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_2261	<b>803.2</b>	Yes	Yes	0.99	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_2264	<b>1022.02</b>	Yes	Yes	0.99	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_2265	<b>514.69</b>	Yes	Yes	0.99	Adhesin-like protein	Secreted	Cytoplasmic	Extracellular	Pectin lyase fold
D5_2266	<b>384.67</b>	Yes	Yes	0.95	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_2335	46.85	Yes	Yes	0.76	Adhesin-like protein	Secreted	Cytoplasmic	Periplasmic	Bacterial Ig-like, domain 1
D5_2337	66.75	Yes	Yes	0.82	Adhesin-like protein	Secreted	Extracellular	Extracellular	Glycoside hydrolase superfamily
D5_2354	44.76	No	No	0.72	Adhesin-like protein	Secreted	Extracellular	Extracellular	Peptidase C1A
D5_2356	11.62	No	No	0.77	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_2357	18.51	No	No	0.90	Adhesin-like protein	Cytoplasm	Cytoplasmic	Extracellular	Pectin lyase fold
D5_2407	131.47	No	No	0.70	Adhesin-like protein	Cytoplasm	Extracellular	Extracellular	Peptidase C1A

<sup>1</sup>. kDa: Predicted molecular weight in kilodaltons, proteins above 200 kilodaltons are displayed in bold. <sup>2</sup>. TMH: Transmembrane helices.

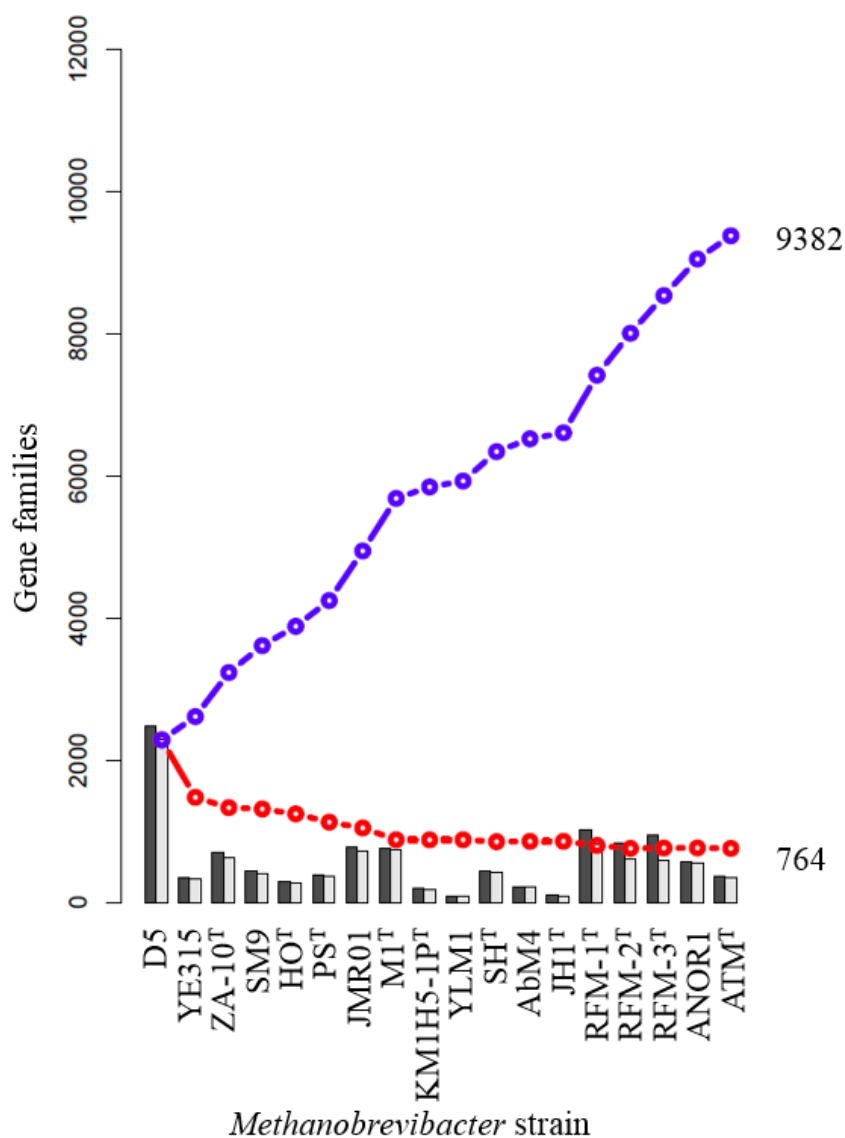
*Synten*y. The synteny between completed genomes was analysed using D5 as a reference (Figure 6.13). D5 shares a high degree of synteny with YE315 and SM9, the syntenic breakpoint in SM9 included a region downstream of an adhesin-like protein (D5\_0186) and a region upstream of a hypothetical protein (D5\_2337), encompassing approximately 2.3 Mb (Figure 6.13E). The syntenic breakpoint in YE315 included an acetyl-CoA acetyltransferase gene (D5\_0873) and a region downstream of a non-ribosomal peptide synthase (D5\_1744), and encompasses approximately 1 Mb (Figure 6.13F), whereas a weaker X-shaped alignment was observed between D5 and AbM4 with multiple syntenic breakpoints (Figure 6.13A).



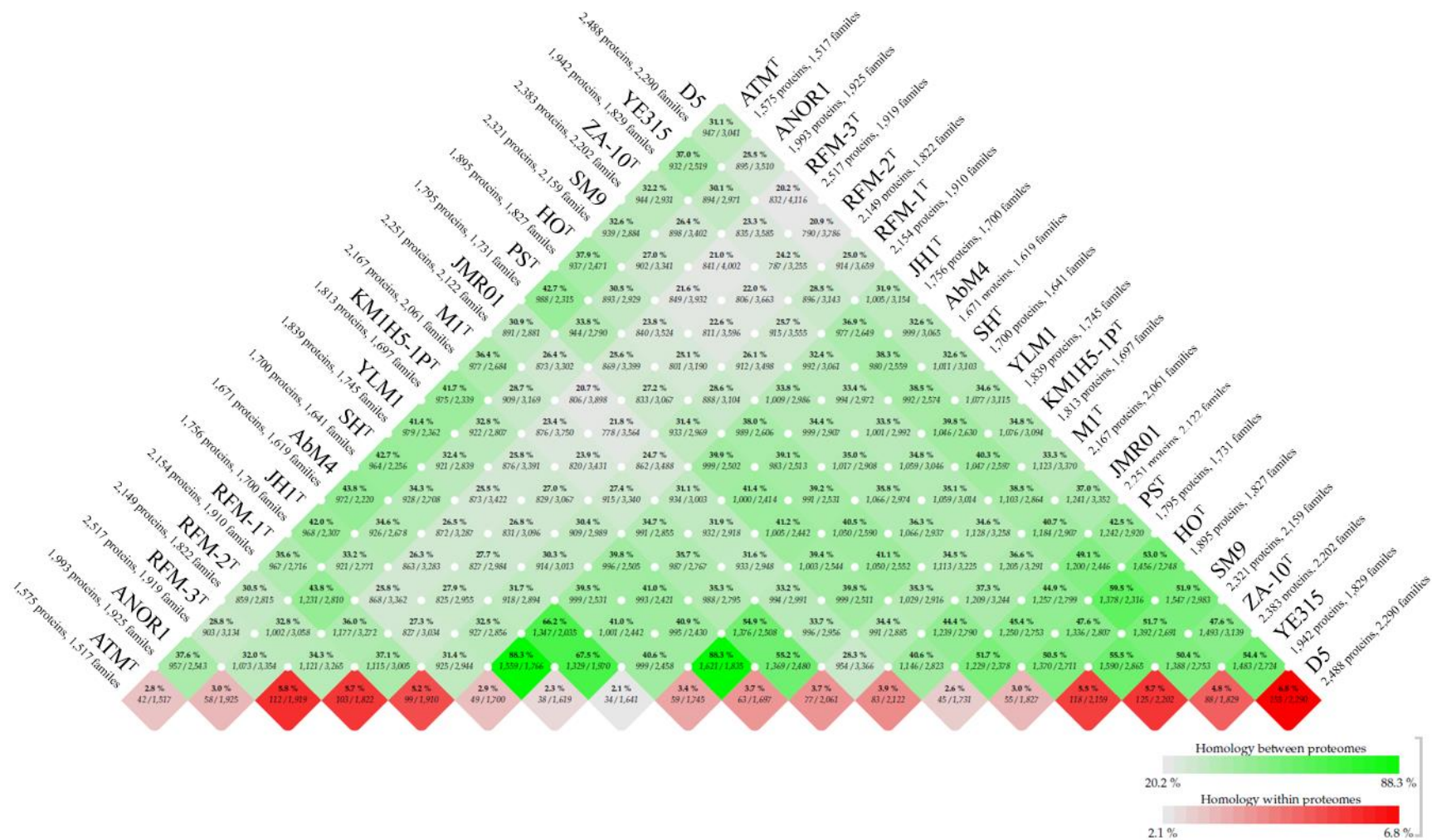
**Figure 6.13 Gene synteny plots for completed genomes of *Methanobrevibacter* spp.** PRoMmer alignments of the D5 genome against genomes from *Methanobrevibacter* spp. are shown. The alignments were plotted using MUMmer (Delcher *et al.* 2003) with forward matches shown in red and reverse matches in blue. The units displayed on both axes are in Mb.

### 6.2.5 Comparative analysis of gene families in *Methanobrevibacter* spp.

The term core-genome was used to describe the genes or gene families present in all *Methanobrevibacter* genomes analysed while the pan-genome was used to define the full complement of genes or gene families present in all sequenced *Methanobrevibacter* genomes. The core genome consists of 764 gene families and the pan-genome consists of 9,382 gene families (Figure 6.14).



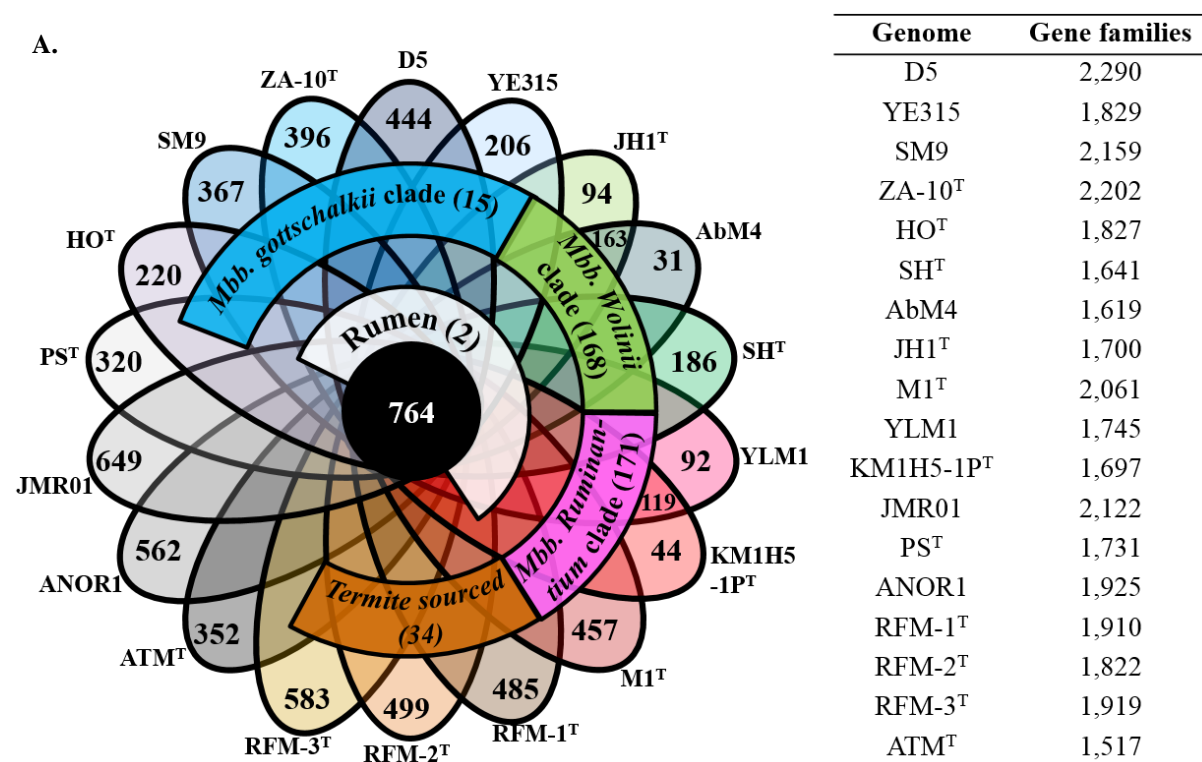
**Figure 6.14 Pan-genome and core-genome plot of *Methanobrevibacter* genomes.** The blue line indicates the cumulative pan-genome, and the red line indicates the core-genome. The black bars indicate the number of new genes with the addition of each genome across the x-axis; the light grey bars indicates new gene families.



**Figure 6.15 BLASTP matrix illustrating the number of conserved protein families between *Methanobrevibacter* genomes.** Conservation is defined as 50% coverage and 50% identity. The color intensities are based on the relative percentage of conserved gene families; green depicts conserved protein families between genomes, red depicts protein homology within a genome.

The pattern of gene family conservation as illustrated in Figure 6.15 shows SH<sup>T</sup>, AbM4 and JH1<sup>T</sup> are similar to each other, while KM1H5-1P<sup>T</sup> and YLM1 are more similar to each other. The analysis shows that D5 has the highest percentage of paralogs. The highest predicted gene family conservation occurs between the JH1<sup>T</sup>/AbM4 pair and between the YLM1/KM1H5-1P<sup>T</sup> pair at 88.3% (Figure 6.15).

The number of gene families in the core-genome corresponds with the phylogenetic relationship between the organisms, as does the number of gene families specific to each genome. The number of gene families specific to JMR01 is high, with 649 (30.6%) of its total gene families being specific to its genome (Figure 6.16A).



**Figure 6.16 Conserved and novel gene families among the 18 *Methanobrevibacter* genomes.** A. Venn diagram indicating the number of conserved gene families between genomes based on BLASTp analysis, using a 50% identity and 50% coverage cutoff. Regions that do not overlap with other genomes depict the number of unique gene families under the same criteria. The conserved gene families within each clade and conserved gene families between all genomes of rumen origin are presented in brackets. B. Table listing the total number of gene families in each of the *Methanobrevibacter* genomes.

*Analysis of the Methanobrevibacter core genome.* The number of gene families in the core genome is summarised in Figure 6.17 and Table 6.14. A full list of core genome gene families is presented in the Appendix (Table A.6.5). The core genome lacks the tryptophan and proline biosynthesis operons. The JMR01 genome is predicted to lack tryptophan biosynthesis genes, and all *Methanobrevibacter* genomes are predicted to lack proline biosynthesis genes.

**Table 6.14 Conserved and novel genes of *Methanobrevibacter* genomes by functional categories**

COG	Code	Gene families						D5 novel
		D5 Total	Core	Rumen	<i>Mbb. gottschalkii</i> clade	<i>Mbb. ruminantium</i> clade*	<i>Mbb. wolinii</i> clade*	
Translation	J	149	118	0	0	1	0	1
RNA processing and modification	A	0	0	0	0	0	0	0
Transcription	K	106	32	0	0	3	3	5
Replication, recombination and repair	L	124	37	0	0	5	4	14
Chromatin structure and dynamics	B	2	1	0	0	0	0	0
Cell cycle control, mitosis and meiosis	D	10	4	0	0	0	1	4
Nuclear structure	Y	0	0	0	0	0	0	0
Defense mechanisms	V	39	2	0	0	1	3	9
Signal transduction mechanisms	T	26	2	0	0	0	3	4
Cell wall/membrane biogenesis	M	106	24	1	1	3	4	26
Cell motility	N	4	0	0	0	0	0	0
Cytoskeleton	Z	0	0	0	0	0	0	0
Extracellular structures	W	0	0	0	0	0	0	0
Intracellular trafficking and secretion	U	11	7	0	0	0	0	0
Post-translational modification, protein turnover, chaperones	O	66	23	0	0	4	3	12
Energy production and conversion	C	173	85	1	1	4	4	9
Carbohydrate transport and metabolism	G	51	27	0	1	0	1	2
Amino acid transport and metabolism	E	126	64	0	0	4	9	7
Nucleotide transport and metabolism	F	51	35	0	0	0	2	0
Coenzyme transport and metabolism	H	116	56	0	0	17	3	2
Lipid transport and metabolism	I	32	10	0	1	1	1	5
Inorganic ion transport and metabolism	P	59	21	0	0	3	7	2
Secondary metabolites biosynthesis, transport and catabolism	Q	23	2	0	0	0	0	2
General function prediction only	R	262	90	0	1	16	29	28
Function unknown	S	223	87	0	2	11	16	23
Unclassified COG	Unclassified COG	283	21	0	4	32	33	74
Not in COGs	-	445	16	0	4	66	42	215

\*The gene families in the *Mbb. ruminantium* clade used M1<sup>T</sup> as reference for COG identification, and the gene families in the *Mbb. wolinii* clade used AbM4 as reference.







**Table 6.16 Gene families uniquely conserved in the *Mbb. gottschalkii* clade by functional category**

Locus_tag	Predicted gene product	COG category
D5_0388	hypothetical protein	[unclassified]
D5_0628	conserved hypothetical transmembrane protein	[unclassified]
D5_0738	formate dehydrogenase alpha subunit, flpA	[C]
D5_0824	NADPH-dependent FMN reductase	[R]
D5_0940	hypothetical transmembrane protein	[S]
D5_1099	ATP-grasp domain containing protein	[I]
D5_1317	hypothetical transmembrane protein	[unclassified]
D5_1344	conserved hypothetical transmembrane protein	[unclassified]
D5_1640	hypothetical protein	[M]
D5_1755	hypothetical protein	Not in COGs
D5_1759	MFS transporter	[G]
D5_1853	hypothetical protein	Not in COGs
D5_1895	hypothetical protein	Not in COGs
D5_1990	hypothetical protein	[S]
D5_2064	hypothetical protein	Not in COGs

\*only one gene from D5 for each gene family is represented

*Uniquely conserved gene families in the genomes of the Mbb. ruminantium clade.* Amongst the 171 gene families conserved within the *Mbb. ruminantium* clade, 125 are either classified as poorly characterised or not in COGs, and 17 are classified under coenzyme transport and metabolism (Table 6.14, Table A.6.6), including genes encoding a 6-carboxyhexanoate—CoA ligase BioW (MRU\_RS10255, 2595159567, YLM1\_1498), 8-amino-7-oxonanoate synthase BioF (MRU\_RS10260, 2595159566, YLM1\_1499) and adenosylmethionine--8-amino-7-oxonanoate aminotransferase BioA (MRU\_RS10475, 2595159351, YLM1\_1816), which are involved in biotin biosynthesis using pimelate as a precursor.

*Uniquely conserved gene families in the genomes of the Mbb. wolinii clade.* Amongst the 168 gene families conserved within the *Mbb. wolinii* clade, 120 are either classified as poorly characterised or not in COGs, and nine gene families are classified under amino acid transport and metabolism (Table 6.14, Table A.6.7), including five genes involved in tryptophan biosynthesis *trpACDEF*. Despite 17 *Methanobrevibacter* genomes possessing a complete tryptophan biosynthesis operon, the *trpACDEF* genes in members of *Mbb. wolinii* clade are different to the *trpACDEF* genes in 14 other *Methanobrevibacter* genomes analysed. Using *trpC* and *trpD* genes as examples, *trpC* gene has an average of 41.4% aa identity between *Mbb. wolinii* clade and other *Methanobrevibacter* spp., *trpD* gene has an average of 39.1% aa identity between *Mbb. wolinii* clade and other *Methanobrevibacter* spp., whereas the *trpC* genes and *trpD* genes are similar when compared within *Mbb. wolinii* clade (79.7%, 84% aa respectively) and compared within members outside of the *Mbb. wolinii* clade (67%, 71% aa respectively). In members of *Mbb. wolinii* clade, *trpCD* genes shares a high homology to members of the genera *Methanosphaera* and *Methanobacterium*, while *trpCD* genes in other *Methanobrevibacter* genomes share a high homology to members of the phylum *Firmicutes* and the genus *Clostridium*. Four gene families are classified under cell

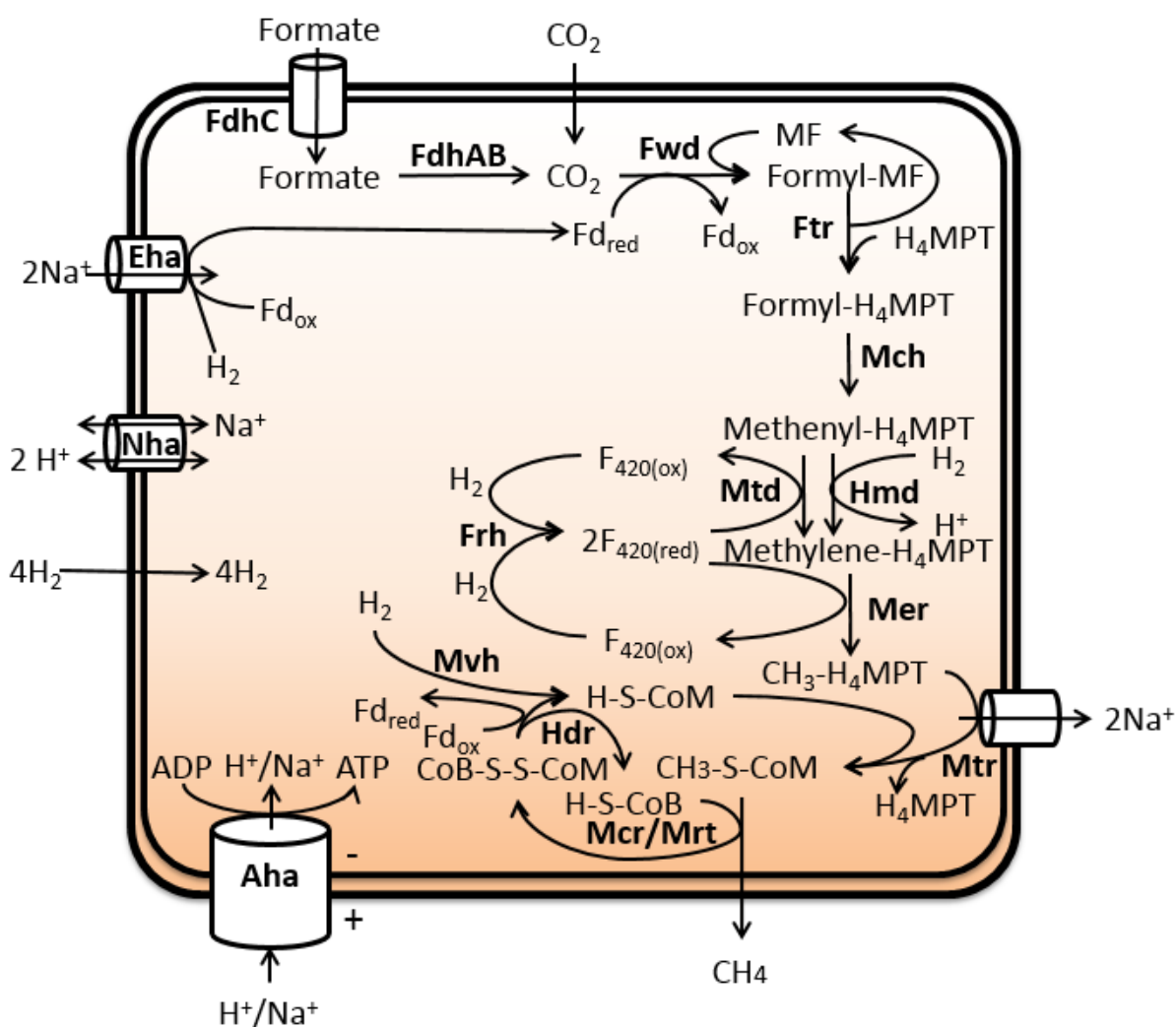
wall/membrane/envelope biogenesis, including a gene encoding aspartate racemase (2540853571, 2558933547, 2553938151) involved in conversion of L-aspartate to D-aspartate, this gene not predicted outside of *Mbb. wolinii* clade.

*Novel gene families in the Methanobrevibacter sp. D5 genome.* The analysis of gene family conservation identified 444 unique gene families in *Methanobrevibacter* sp. D5 (Table A.6.8), of which, 342 are classified as poorly characterised, 26 genes are classified as cell wall/membrane/envelope biogenesis genes, including six glycosyl transferase genes (D5\_0223, D5\_0462, D5\_0463, D5\_0484, D5\_0488, D5\_1305) that may be involved in exopolysaccharide formation and four adhesin-like protein encoding genes (D5\_0459, D5\_0965, D5\_1473, D5\_2175). In addition, 46 other adhesin-like protein encoding genes were also identified as unique. Two genes predicted to be an endonuclease (D5\_2164) and a methylase (D5\_2165) of a type III restriction system are considered novel under the criteria used in this study. A non-ribosomal surfactin synthetase (D5\_0482) was also identified as unique.

## 6.2.6 Metabolic pathway construction of *Methanobrevibacter* sp. D5

D5 genes with a predicted metabolic function are listed in Table A.6.1. A total of 1,232 genes, representing 49.5% of the genome, have been assigned to a metabolic pathway.

### Energy metabolism



**Figure 6.18 Hydrogenotrophic methanogenesis pathway of *Methanobrevibacter* sp. D5.** Enzyme names are displayed in bold. The enzymes involved are: formate transporter (FdhC), formate dehydrogenase (FdhAB), formylmethanofuran dehydrogenase (Fwd), formylmethanofuran-tetrahydromethanopterin formyl transferase (Ftr), N<sup>5</sup>, N<sup>10</sup>-methenyl-H<sub>4</sub>MPT cyclohydrolase (Mch), F<sub>420</sub>-dependent methylene-H<sub>4</sub>MPT dehydrogenase (Mtd), F<sub>420</sub> reducing hydrogenase (Frh), methylene-H<sub>4</sub>MPT reductase (Mer), tetrahydromethanopterin S-methyltransferase (Mtr), H<sub>2</sub>-dependent methylene-H<sub>4</sub>MPT dehydrogenase (Hmd), Methyl-CoM reductase (Mcr/Mrt), heterodisulfide reductase (Hdr), methyl-viologen hydrogenase (Mvh). The cofactors involved are: methanofuran (MF), ferredoxin (Fd), tetrahydromethanopterin (H<sub>4</sub>MPT), coenzyme M (CoM), coenzyme B (CoB), and cofactor F<sub>420</sub> (F<sub>420</sub>). Components involved in membrane potential generation and energy generation are energy conserving hydrogenase (Eha), Na<sup>+</sup>/H<sup>+</sup> antiporter (Nha), tetrahydromethanopterin S-methyltransferase (Mtr) and A<sub>1</sub>A<sub>0</sub> ATP synthase (Aha).

*Methanogenesis.* Based on the genome sequence, methanogenesis is the sole energy generating mechanism of *Methanobrevibacter* sp. D5. The genes predicted to be involved in methanogenesis are summarized in Figure 6.18 and Table A.6.1. The D5 genome encodes genes required for energy generation via hydrogenotrophic methanogenesis (Table A.6.1), including the formate transporter FdhC (D5\_2393), formate dehydrogenase FdhAB (D5\_2391, D5\_2392), FlpAB (D5\_0738, D5\_0739) and its accessory protein FdhD (D5\_02394) and FlpD (D5\_0772), suggesting D5 can utilise carbon dioxide (CO<sub>2</sub>), hydrogen (H<sub>2</sub>) and formate as substrates (Figure 6.18).

The H<sub>2</sub> dependent fixation of CO<sub>2</sub> is carried out by a tungsten containing formylmethanofuran dehydrogenase Fwd, encoded by *fwdHFGDBAC* (D5\_2382 – D5\_2388) and *fwdE* (D5\_1080, D5\_1751) genes. Fwd binds CO<sub>2</sub> to a methanofuran (MF) and reduce it to formyl-MF using the reducing potential supplied from the membrane bound energy conserving [NiFe]-hydrogenase EhaA (D5\_1770 – D5\_1787) with H<sub>2</sub> via ferredoxin (Figure 6.18). The formyl group is then transferred to the tetrahydromethanopterin (H<sub>4</sub>MPT) by formylmethanofuran-tetrahydromethanopterin formyl transferase Ftr (D5\_0210, D5\_1769), producing formyl-H<sub>4</sub>MPT. The formyl-H<sub>4</sub>MPT is reversibly catalysed to N<sup>5</sup>, N<sup>10</sup>-methenyl-H<sub>4</sub>MPT by N<sup>5</sup>, N<sup>10</sup>-methenyl-H<sub>4</sub>MPT cyclohydrolase Mch (D5\_1626). Methenyl-H<sub>4</sub>MPT is reduced to Methylene-H<sub>4</sub>MPT by either F<sub>420</sub>-dependent methylene-H<sub>4</sub>MPT dehydrogenase Mtd (D5\_0079, D5\_0741) or H<sub>2</sub>-dependent methylene-H<sub>4</sub>MPT dehydrogenase Hmd (D5\_2132). Methylene-H<sub>4</sub>MPT is subsequently reduced to methyl-H<sub>4</sub>MPT by Methylene-H<sub>4</sub>MPT reductase Mer (D5\_2110), this reaction is reversible and F<sub>420</sub> dependent. The F<sub>420</sub> hydrogenase Frh encoded by *frhADGB* (D5\_0174 – D5\_0177) is responsible for reducing F<sub>420</sub> using H<sub>2</sub>, supplying reducing potential for methanogenesis. The methyl group is transferred from methyl-H<sub>4</sub>MPT to coenzyme M (CoM) via tetrahydromethanopterin S-methyltransferase complex Mtr (D5\_0289 – D5\_0296), producing methyl-CoM. Mtr is coupled to the generation of a Na<sup>+</sup> gradient across the cytoplasmic membrane, and is one of the enzyme complexes responsible for the generation of a membrane potential. An additional copy of *mtrA* (D5\_0510) and *mtrH* (D5\_0350) genes were identified. The reduction of methyl-coenzyme M to methane is carried out by the methyl-coenzyme M reductase enzyme complex; two isoenzymes are present in the D5 genome, a methyl-CoM reductase I Mcr (D5\_0284 – D5\_0288) and a methyl-CoM reductase II Mrt (D5\_0742 – D5\_0745).

*Methanogenesis cofactors.* In the hydrogenotrophic methanogenesis pathway, three cofactors act as the C<sub>1</sub> moiety is transferred sequentially from CO<sub>2</sub> to MF, H<sub>4</sub>MPT and CoM, and reduced

to CH<sub>4</sub>. The cofactors involved in the last step of methanogenesis are the same as described in ISO4-H5, where the cofactor F<sub>430</sub> acts as the nickel-porphinoid group central to Mcr/Mrt catalytic activity, and coenzyme B (CoB) is used by the Mcr and Mrt complex for methyl-coenzyme M reduction to produce methane (Figure 6.18). This process utilises the reducing potential carried by HS-CoB and generates CoM-S-S-CoB (heterodisulfide). Both cofactors need to be recycled into an active state, and this is carried out by a heterodisulfide reductase (Hdr). D5 encodes two sets of Hdr genes *hdrABC* (D5\_0649 - 0650, D5\_1502 - 1504, D5\_2467) with an additional copy of *hdrB* (D5\_1005) gene. D5 encodes the methyl viologen hydrogenase Mvh required for coupling H<sub>2</sub> to provide reducing potential for the reducing CoM-S-S-CoB and ferredoxin via electron bifurcation MvhDGAB (D5\_0302 - 0305) (Figure 6.18). A second cofactor involved in the coupling of H<sub>2</sub> to provide reducing potential is cofactor F<sub>420</sub>, Mtd and Mer depend on F<sub>420</sub>H<sub>2</sub> as the source of reductant.

*Generation of membrane potential from methanogenesis and potential sources of reducing potential driving methanogenesis.* ATP synthase in *Methanobrevibacter* spp. can be driven by either Na<sup>+</sup> or H<sup>+</sup> gradient. Methanogenesis is coupled to the generation of a Na<sup>+</sup> gradient, the conversion of methyl-H<sub>4</sub>MPT to methyl-CoM by Mtr (D5\_0289 - 0296) is coupled to the translocation of Na<sup>+</sup> across the cell membrane, thereby generating a Na<sup>+</sup> gradient. The first step of hydrogenotrophic methanogenesis is carried out by a Fwd, which utilises the reducing potential supplied from reduced ferredoxins (Fd<sub>red</sub>), the reduction of ferredoxin is catalysed by membrane associated energy conserving hydrogenase A complex (EhaA) (D5\_1770 - D5\_1787) driven by a Na<sup>+</sup> motive force in the hydrogenotrophic methanogens. A second energy conserving hydrogenase complex (EhaB) (D5\_0225 - 0241) was identified in the D5 genome, which is associated with regeneration of electron carriers involved in autotrophic CO<sub>2</sub> fixation and acetate assimilation. Na<sup>+</sup>/H<sup>+</sup> antiporters (Nha) (D5\_0516, D5\_1713) are also involved in ion gradient formation, they are responsible for coupling and cycling of Na<sup>+</sup>, H<sup>+</sup> and pH homeostasis. Nha could compete with the ATP synthase for Na<sup>+</sup>. Frh and F<sub>420</sub>H<sub>2</sub> supplies reducing potential to methanogenesis independently of membrane potential. In addition, a gene encoding NADP-dependent F<sub>420</sub> reductase, NpdG (D5\_1461), has been predicted in the D5 genome, as well as two alcohol dehydrogenase genes (D5\_1302, D5\_1857) that may work in conjunction with NpdG to supply reducing potential.

*Methanogenesis marker proteins.* Homologues of methanogenesis marker proteins 1 through 17 are present within the D5 and all of the *Methanobrevibacter* spp. genomes analysed.

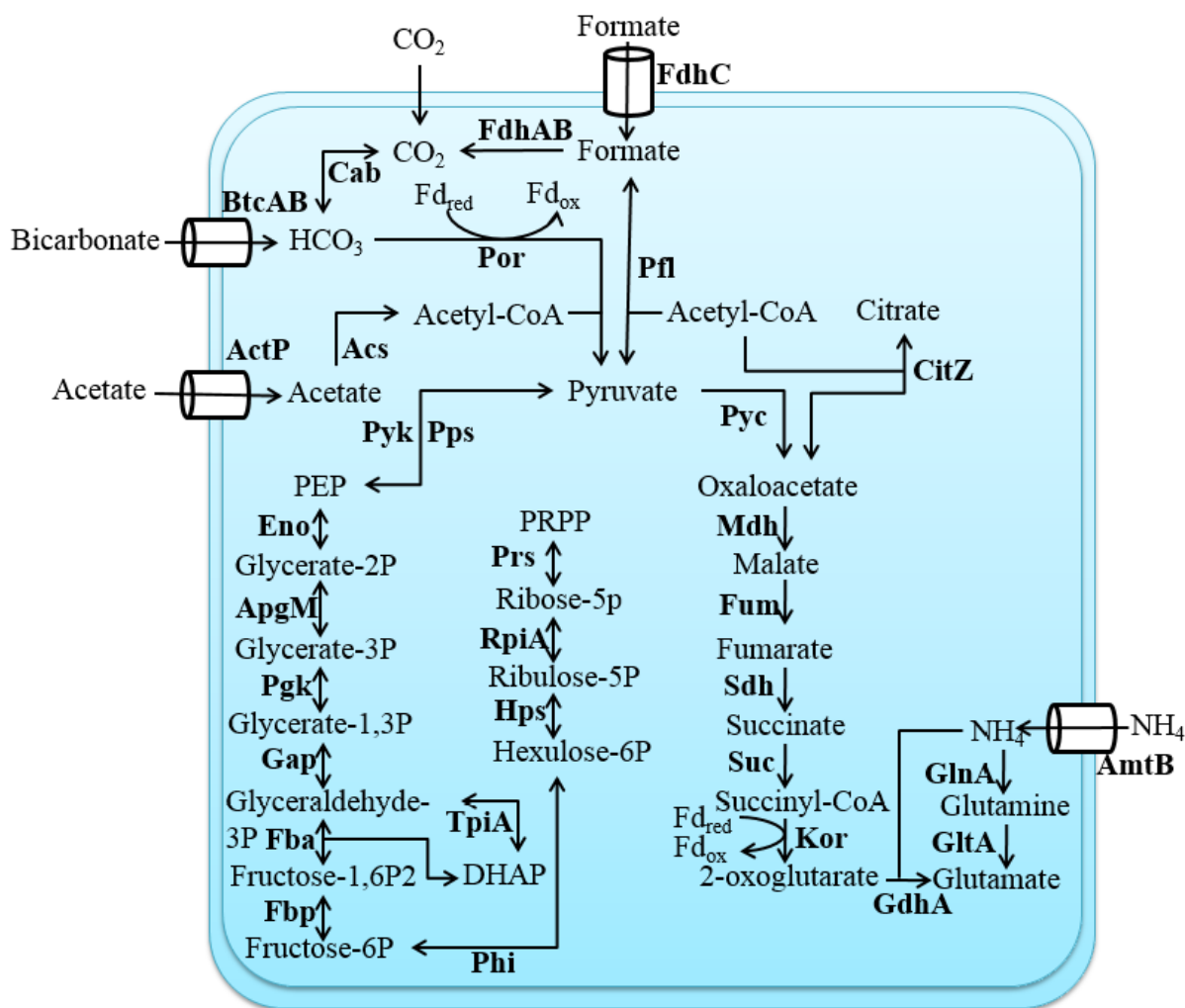
### *Central carbon metabolism*

D5 is predicted to encode the genes required for a partial TCA cycle, gluconeogenesis and the pentose phosphate pathway.

*One-carbon metabolism.* D5 has a complement of genes allowing it to utilise acetate, bicarbonate, formate and CO<sub>2</sub> as a potential carbon source (Figure 6.09). D5 encodes two acetyl-CoA synthetase, Acs, a SSS family solute/Na<sup>+</sup> symporter homologous to an acetate permease, ActP, as well as a holoenzyme of pyruvate synthase, PorABCD (Table A.6.01) that allows the utilisation of acetate and the subsequent production of acetyl-CoA and pyruvate. Por utilises a reduced ferredoxin as the source of its reducing potential.

Formate is involved in hydrogenotrophic methanogenesis as well as central carbon metabolism. D5 encodes a pyruvate formate-lyase, Pfl, and the associated Pfl-activating protein, allowing the conversion of formate to pyruvate. In addition, D5 is predicted to possess a set of bicarbonate ABC transporter genes, *btcABC*, allowing bicarbonate to be utilised as a carbon source by Por, or interconverted to CO<sub>2</sub> by carbonic anhydrase, Cab.

*TCA cycle.* The D5 genome is predicted to have an incomplete reductive TCA cycle from oxaloacetate to 2-oxoglutarate (Figure 6.19, Table A.6.1). D5 has the 2-oxoglutarate ferredoxin oxidoreductase *korDABC* genes required to convert succinyl-CoA to 2-oxoglutarate, this reaction utilises reduced ferredoxin as the source of its reducing potential. The ferredoxins associated with Kor and Por are reduced by the energy conserving hydrogenase B (EchB).



**Figure 6.19 Central carbon metabolism of *Methanobrevibacter* sp. D5.** Enzymes are displayed in bold. The enzymes involved are: bicarbonate ABC transporter (BtcAB), carbonic anhydrase (Cab), formate/nitrite transporter (FdhC), formate dehydrogenase (FdhAB), ferredoxin (Fd), pyruvate ferredoxin oxidoreductase (Por), pyruvate formate-lyase (Pfl), cation/acetate symporter (ActP), acetyl-CoA synthetase (Acs), pyruvate kinase (Pyk), Phosphoenolpyruvate synthase (Pps), enolase (Eno), 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (ApgM), phosphoglycerate kinase (Pgk), glyceraldehyde3-phosphate dehydrogenase (Gap), fructose-bisphosphate aldolase (Fba), triose-phosphate isomerase (TpiA), fructose 1,6-bisphosphatase (Fbp), 3-hexulose-6-phosphate isomerase (Phi), 3-hexulose-6-phosphate synthase (Hps), ribose 5-phosphate isomerase A (RpiA), ribose-phosphate diphosphokinase (Prs), pyruvate carboxylase (Pyc), malate dehydrogenase (Mdh), citrate synthase (CitZ), fumarate hydratase (Fum), succinate dehydrogenase (Sdh), succinyl-CoA synthetase (Suc), 2-oxoglutarate ferredoxin oxidoreductase (Kor), glutamate dehydrogenase (GdhA), glutamine synthase (GlnA), glutamate synthase (GltA) and ammonium transporter (AmtB). The compounds involved are phosphoenolpyruvate (PEP), dihydroxyacetone phosphate (DHAP) and phosphoribosyl pyrophosphate (PRPP).

**Gluconeogenesis.** D5 encodes all of the genes required for gluconeogenesis from pyruvate to fructose-6-phosphate (Figure 6.19, Table A.6.1).

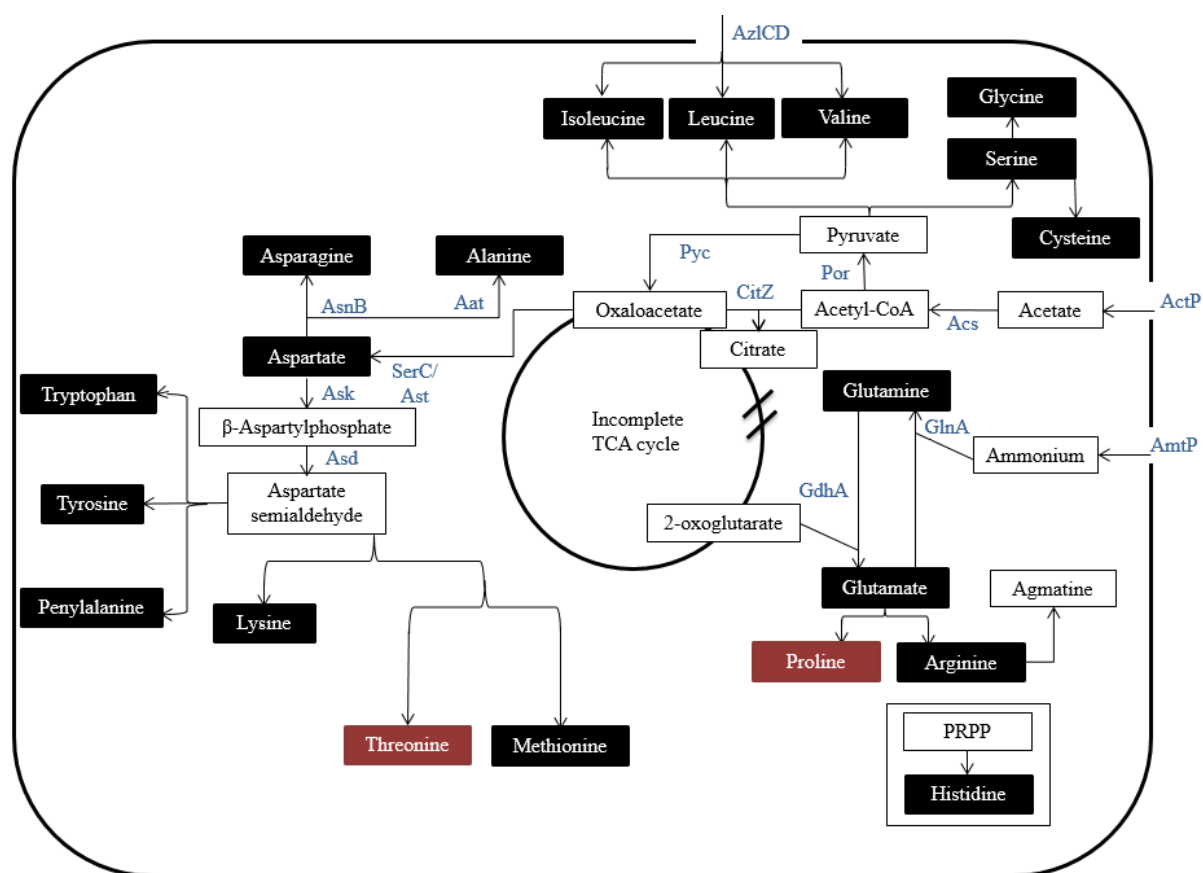
**Pentose phosphate pathway.** The D5 genome encodes all the genes required for a non-oxidative pentose phosphate pathway and phosphoribosyl pyrophosphate (PRPP) production (Figure 6.19, Table A.6.1). Two homologues of the 3-hexulose-6-phosphate isomerase (Phi) were found.

## Amino acid biosynthesis

The D5 is predicted to be able to make 18 amino acids based on the genes present in the genome and the genes involved are listed (Figure 6.20, Table A.6.1). D5 is predicted to be unable to produce proline and threonine. D5 encodes an amino acid carrier protein (D5\_1090) and an amino acid ABC transporter (D5\_1294 – 1296) that may be able to facilitate uptake of these amino acids to overcome these auxotrophies.

**Proline.** The genes required for proline biosynthesis, *proABC*, were not found in the D5 genome.

**Threonine.** Threonine is produced from aspartate via a homoserine intermediate. The D5 genome is predicted to lack the *thrB* gene, which encodes the homoserine kinase required to convert homoserine to *O*-phospho-L-homoserine, the substrate of threonine synthase, ThrC (D5\_1657). Therefore, D5 is predicted to be incapable of threonine biosynthesis.



**Figure 6.20 Amino acid biosynthesis of *Methanobrevibacter* sp. D5.** Amino acids predicted to be synthesized are displayed in black text boxes. Amino acids predicted not to be synthesized are displayed in red text boxes. Intermediates and molecules are displayed in white text box. The enzymes involved are displayed in blue text, including citrate synthase (CitZ), pyruvate:ferredoxin oxidoreducta (Por), acetyl-CoA synthase (Acs), pyruvate



carboxylase (Pyc), glutamine synthetase (GlnA), : glutamate dehydrogenase (GdhA), ammonium transporter (AmtP), cation/acetate symporter (ActP), branched chain amino acid transporter (AzlCD), asparagine synthase (AsnB), alanine aminotransferase (Aat), broad substrate specificity phosphoserine aminotransferase (SerC), aspartate amino transferase (Ast), aspartate kinase (Asp), aspartate semialdehyde dehydrogenase (Asd). Phosphoribosylpyrophosphate is abbreviated PRPP.

### *Vitamins and cofactors*

*Cofactor F<sub>430</sub>*. As described in Chapter 3, Section 3.2.4, uroporphyrinogen is the prerequisite to tetrapyrrole cofactors and cobalamin. The D5 genome possesses the full complement of genes required to produce uroporphyrinogen (Table A.6.1), including glutamyl-tRNA reductase, HemA (D5\_0334), delta-aminolevulinic acid dehydratase, HemB (D5\_0789), porphobilinogen deaminase, HemC (D5\_0519), uroporphyrinogen III synthase, HemD (D5\_0819) and glutamate-1-semialdehyde-2,1-aminomutase, HemL (D5\_0045). The only identified enzyme in the pathway of cofactor F<sub>430</sub> biosynthesis, uroporphyrin-III C-methyltransferase, CorA (D5\_0865), is also present. No operon arrangements were observed.

*Cobalamin/thiamine*. The sirohydrochlorin intermediate produced from F<sub>430</sub> biosynthesis can be used for cobalamin (vitamin B<sub>12</sub>) biosynthesis (Table A.6.1). The D5 genome is predicted to encode all but 2 genes required for the anaerobic pathway of cobalamin biosynthesis, the missing genes are *cobC* and cobirinic acid-a,c-diamide reductase. D5 is also predicted to encode genes required for thiamine biosynthesis, including two copies of *thiC* and *thiD* genes (Table A.6.1).

*Riboflavin/FMN/FAD*. The D5 genome is predicted to encode the full complement of genes required to produce riboflavin (Table A.6.1).

*MF/H<sub>4</sub>MPT/CoM/CoB/F<sub>420</sub>/F<sub>390</sub>*. There are five important coenzymes involved in methanogenesis, MF, H<sub>4</sub>MPT, CoM, CoB and cofactor F<sub>420</sub>. The D5 genome is predicted to possess the full complement of genes required to produce CoM, CoB and cofactor F<sub>420</sub> (Table A.6.1). D5 is predicted to encode all except the 2-furaldehyde phosphate aminotransferase *mtnC* gene required for MF biosynthesis (Table A.6.1). The D5 genome is also predicted to possess eight genes involved in the biosynthesis of H<sub>4</sub>MPT, however the genes involved in the later part of the biosynthetic pathway remain to be identified. Cofactor F<sub>390</sub> is produced from cofactor F<sub>420</sub> when the methanogen is exposed to oxygen. The D5 genome is predicted to possess two *ftsA* genes encoding coenzyme F<sub>390</sub> synthetase (D5\_1079, D5\_1085).

*Ubiquinone.* The D5 genome encodes five genes predicted to be involved in the ubiquinone biosynthesis, including 3-polyprenyl-4-hydroxybenzoate decarboxylase UbiX, three copies of *ubiB* genes encoding 2-polyprenylphenol 6-hydroxylase and a UbiD family decarboxylase gene, however the function of these genes remain to be elucidated.

*Nicotinate.* The D5 genome has all the genes encoding the enzymes required for nicotinate biosynthesis (Table A.6.1).

*Biotin.* The D5 genome is predicted to contain three genes involved in biotin synthesis, two biotin synthases, BioB (D5\_1332, D5\_2133) and a biotin-acetyl-CoA-carboxylase ligase, BirA (D5\_0677), no others genes required for biotin biosynthesis were identified.

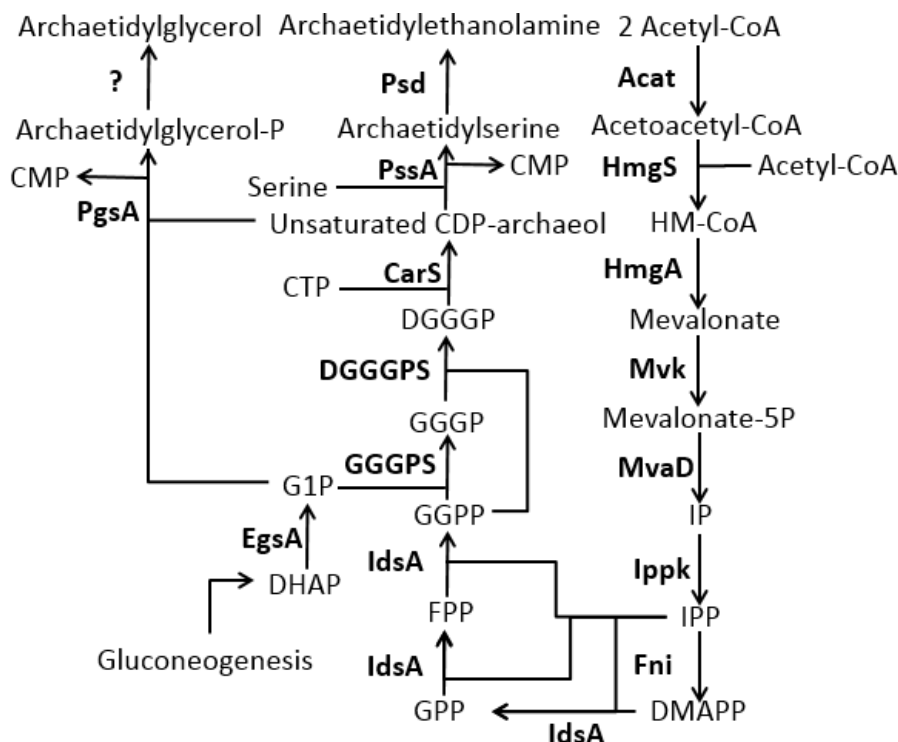
*Glutathione.* The D5 genome is predicted to encode a gamma-glutamylcysteine synthetase *gshA* gene and a bifunctional glutamate-cysteine ligase/glutathione synthetase *gshF* gene required for glutathione biosynthesis, and two copies of glutathione peroxidase *gpxA* genes and glutathione-disulfide reductase *gor* genes (Table A.6.1).

#### *Nitrogen metabolism*

D5 is predicted to encode two copies of the ammonium transporter *amtB* gene (D5\_0957, D5\_1675). Ammonium uptake is linked to glutamine and glutamate biosynthesis by glutamine synthase *glnA* (D5\_2378) and glutamate dehydrogenase *gdhA* (D5\_0429) as depicted in Figure 6.19. The D5 genome also predicts two genes involved in regulation of nitrogen assimilation, a gene encoding the transcriptional repressor adjacent to *glnA*, *nrpR* (D5\_2379), and two nitrogen regulatory protein P-II *glnK* genes (D5\_0958, D5\_1674). The D5 genome is predicted to lack genes encoding nitrogenase, but D5 has several other proteins predicted to be involved in nitrogen assimilation (Table A.6.1), including a nitrogenase cofactor biosynthesis protein, NifB (D5\_0636), a dinitrogenase iron-molybdenum cofactor biosynthesis protein (D5\_2272), a cysteine desulfurase, NifS (D5\_1698), and a hydroxylamine reductase, Hcp (D5\_0749).

#### *Lipids and cell envelope*

The D5 genome has a full complement of genes required to produce archaeol via the mevalonate pathway (Table A.6.1), the proposed *de novo* phospholipid biosynthesis pathway is displayed in Figure 6.21.



**Figure 6.21 Predicted phospholipid biosynthesis pathway in *Methanobrevibacter* sp. D5.** Enzymes are displayed in bold. The enzymes involved are: acetyl-CoA acetyltransferase (Acat), hydroxymethylglutaryl-CoA synthase (HmgS), hydroxymethylglutaryl-CoA reductase (HmgA), mevalonate kinase (Mvk), phosphomevalonate decarboxylase (MvaD), isopentenyl diphosphate kinase (Ippk), isopentenyl-diphosphate  $\delta$ -isomerase (Fni), bifunctional short chain isoprenyl diphosphate synthase (IdsA), NAD(P)-dependent glycerol-1-phosphate dehydrogenase (EgsA), geranylgeranylgeranyl phosphate synthase (GGGPS), digeranylgeranylgeranyl phosphate synthase (DGGGPS), CDP-archaeol synthase (CarS), CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (PgsA), phosphatidylserine synthase (PssA), phosphatidylserine decarboxylase (Psd). The compounds involved are hydroxymethylglutaryl-CoA (HM-CoA), isopentenyl phosphate (IP), isopentenyl diphosphate (IPP), 3,3-dimethylallyldiphosphate (DMAPP), geranyl diphosphate (GPP), farnesyl diphosphate (FPP), geranylgeranyl diphosphate (GGPP), dihydroxyacetone phosphate (DHAP), glycerol-1-phosphate (G1P), geranylgeranylgeranyl phosphate (GGGP), digeranylgeranylgeranyl phosphate (DGGGP), cytidine triphosphate (CTP), cytidine monophosphate (CMP).

The biosynthesis of phospholipid utilises acetyl-CoA and D-glyceraldehyde 3-phosphate from central carbon metabolism. Based on the presence of the CDP-archaeol synthase (*carS*) gene, D5 is predicted to use unsaturated CDP-archaeol as an intermediate in phospholipid biosynthesis. The genes in D5 suggest that CDP-archaeol proceeds towards archaeetidylethanolamine and archaeetidylglycerol biosynthesis.

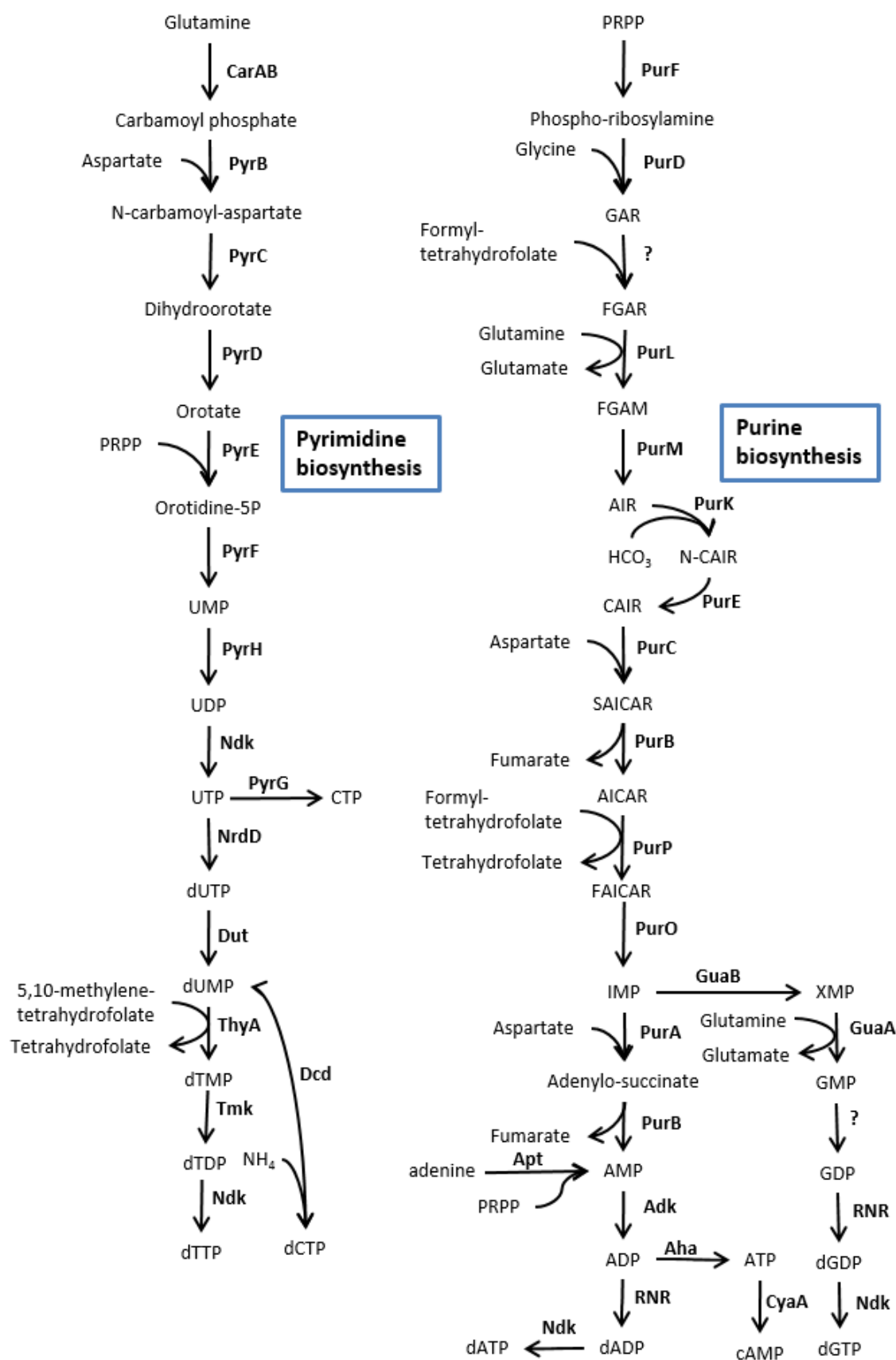
D5 has several genes required to produce the precursor of pseudomurein, UDP-*N*-acetylglucosamine (UDP-GlcNAc) and UDP-*N*-acetylgalactosamine (UDP-GalNAc) from fructose-6-phosphate (Table A.6.1), including two copies each of *glmM* and *glmS*. In addition, several genes are predicted to be associated with pentapeptide formation from UDP-glutamic acid and crosslinking to alanine, lysine and glutamic acid, including six genes belonging to the cell wall biosynthesis protein of Mur ligase family (D5\_0091, D5\_0092, D5\_0520, D5\_0948,

D5\_1536, D5\_1822). The UDP-pentapeptide is linked to a disaccharide by the cell wall biosynthesis protein, phospho-*N*-acetylmuramoyl-pentapeptide-transferase, a cell wall biosynthesis protein UDP-glycosyltransferase and a UDP-*N*-acetylglucosamine diphosphorylase/glucosamine-1-phosphate *N*-acetyltransferase (GlmU) (D5\_1823, D5\_2237, D5\_2238). D5 also has genes involved in teichoic acid and sialic acid biosynthesis (Table A.6.1), which may be crosslinked to the pseudomurein.

### *Nucleic acid metabolism*

The D5 genome is predicted to possess all the genes required for pyrimidine biosynthesis, but two genes are missing to complete the full complement of genes required for *de novo* purine biosynthesis. The proposed pathway of *de novo* biosynthesis is illustrated in Figure 6.22. The two absent genes from purine biosynthesis are phosphoribosylglycinamide formyltransferase (*purT*) gene whose corresponding enzyme converts 5-phospho- $\beta$ -D-ribosyl) glycinamide (GAR) to formyl-(5-phospho- $\beta$ -D-ribosyl)glycinamide (FGAR), and the gene (*gmk*) encoding guanylate kinase that phosphorylates GMP to GDP.

D5 is also predicted to possess genes involved in nucleotide salvage and interconversion (Table A.6.1), including an adenine phosphoribosyltransferase (Apt) gene (D5\_2430), two CMP/dCMP deaminase genes (D5\_1176 and D5\_2402) and a cytidylate kinase (Cmk) gene (D5\_0708).



**Figure 6.22 De novo nucleotide biosynthesis in *Methanobrevibacter* sp. D5.** Enzyme names are displayed in bold. Enzymes involved include carbamoyl phosphate synthase (CarAB), aspartate transcarbamylase (PyrB), dihydroorotase (PyrC), dihydroorotate dehydrogenase (PyrD), orotate phosphoribosyltransferase (PyrE), UMP

kinase (PyrH), broad substrate specificity nucleoside diphosphate kinase (Ndk), anaerobic ribonucleoside-triphosphate reductase (NrdD), CTP synthase (PyrG), dUTP diphosphatase (Dut), thymidylate synthase (ThyA), thymidylate kinase (Tmk), CMP/dCMP deaminase (Dcd), amidophosphoribosyl transferase (PurF), phosphoribosylamineglycine ligase (PurD), N<sup>5</sup>-carboxyaminoimidazole ribonucleotide synthetase (PurK), N<sup>5</sup>-carboxyaminoimidazole ribonucleotide mutase (PurE), phosphoribosylformylglycinamide synthetase (PurL), phosphoribosylformylglycinamide cyclo-ligase (PurM), phosphoribosylaminoimidazole-succinocarboxamide synthase (PurC), bifunctional enzyme adenosuccinate lyase (PurB), bifunctional AICAR transformylase/IMP cyclohydrolase (PurH), adenylosuccinate synthetase (PurA), adenylate kinase (Adk), A<sub>1</sub>A<sub>0</sub> ATP synthase (Aha), GMP synthase (GuaAab), IMP dehydrogenase (GuaB), adenine phosphoribosyltransferase (Apt), adenylate kinase (Adk), adenylate cyclase (CyaA), ribonucleotide-diphosphate reductase (RNR). The compounds involved includes phosphoribose pyrophosphate (PRPP), (5-phospho-β-D-ribose)glycinamide (GAR), formyl-(5-phospho-β-D-ribose)glycinamide (FGAR), formamido-5-phospho-β-D-ribose)acetamidine (FGAM), amino-(5-phospho-ribose)imidazole (AIR), N<sup>5</sup>-carboxyaminoimidazole ribonucleotide (*N*-CAIR), amino-(5-phospho-ribose)imidazole-4-carboxylate (CAIR), 5-phosphoribosyl-4-(*N*-succinocarboxamide)-5-aminoimidazole (SAICAR), amino-1-(5-phosphoribosyl)imidazole-4-carboxamide (AICAR). The nucleotides are abbreviated as dYZP: d: deoxy, Y-[A:adenine, C: cytidine, U: uridine, T: thymidine, G: guanine, I: inosine, X: xanthosine], ZP-[MP: monophosphate, DP: diphosphate, TP: triphosphate], cAMP: cyclic AMP.

### *Cell cycle*

The archaeal cell replication process has been described in detail in Chapter 3, Section 3.2.4. The D5 genome is predicted to possess a full complement of genes required for cell replication, encoding two origin of replication *orc1/cdc6* genes (D5\_0001, D5\_2279) as described previously in Section 6.2.3.

The replication machinery of D5 includes MCM helicase (D5\_2038) that unwinds DNA at the origin, replication factor A (D5\_2470) and single-stranded DNA binding protein (D5\_0815) which stabilizes single stranded DNA. An OB fold nucleic acid binding domain-containing protein (D5\_2250) with unknown function may also take part in stabilization of single stranded DNA. Primase initiates DNA synthesis by producing short primers, and two DNA primase genes have been predicted in the D5 genome; DNA primases PriAB (D5\_1490, D5\_1495) and DnaG (D5\_1934). Primers are extended by DNA polymerase family B, PolB (D5\_0261, D5\_0803) on the leading strand, and DNA polymerase family D, PolD (D5\_2404, D5\_2507) on the lagging strand. The replication requires a sliding clamp PCNA family protein, Pcn (D5\_0164), with assistance of clamp loading replication factor C (D5\_0103, D5\_0104). The sliding clamp also assists other enzymes to process and join the Okazaki fragments, including flap endonuclease Fen (D5\_2373), ribonuclease HII, RnhB (D5\_0320), and ATP-dependent DNA ligase, DnlI (D5\_2246). Ribonuclease HI (D5\_1611) degrades the DNA/RNA hybrid. The DNA integrase/recombinase (D5\_2195) homologous to XerD could be involved in replication termination.

The D5 genome is predicted to encode a gene containing a SMC (structural maintenance of chromosome) domain, and a DNA double-strand break repair protein (Rad50) gene (D5\_1538) that may be involved in genome segregation. D5 is also predicted to encode DNA topoisomerases I (D5\_2361) and VI (D5\_0347, D5\_0348) that are involved in removing DNA supercoiling from single and double stranded DNA, respectively.

The D5 genome is predicted to encode a tubulin-like cell division protein FtsZ (D5\_2227), which carries out cell division by undergoing GTP-dependent polymerization into a Z-ring, with assistance of other proteins, such as the cell division ATPase, MinD (D5\_0031) and ParA/MinD-like ATPase (D5\_1415), and cell division protein pelota PelA (D5\_2239), the Z-ring then drives the cell division. A DNA recombinase gene (D5\_2189) homologous to XerD may be involved in termination of replication.

#### *DNA Recombination and repair*

The D5 genome contains several genes predicted to encode DNA recombination and repair systems (Table A.6.1). The archaeal nucleosomes are an irregular mass of genomic DNA, and requires DNA-binding histones to bind and protect the chromosome, regulating gene expressions. Two archaeal histones are predicted in the D5 genome (D5\_0005, D5\_0593). Histone acetylation and methylation regulates gene expression by controlling the accessibility of the chromosome; a histone acetyltransferase of ELP3 family has been predicted in the D5 genome (D5\_0595). In addition, the DNA binding protein, AlbA (D5\_0027), and the associated silencing NAD-dependent deacetylase, Sir2 family protein (D5\_0215) have been predicted in the D5 genome.

#### *Protein synthesis*

*Transcription.* The D5 genome has a full complement of genes encoding the transcriptional machinery (Table A.6.1). This includes the TATA-box binding protein (D5\_2364), transcription initiation factor, TFIIB (D5\_1931, D5\_2109), the TFIIE alpha subunit (D5\_2230) involved in transcription initiation, the transcription factor S, Tfs (D5\_2436), that enhances fidelity, and the transcription elongation factor, NusA-like protein (D5\_0415), involved in transcriptional termination. D5 is predicted to possess 12 genes encoding RNA polymerase subunits responsible for transcribing the genetic information onto mRNA (Table A.6.1).

*Maturation of tRNA.* The tRNA precursor molecules that are transcribed contain 5' and 3' extensions which need to be removed to produce functional tRNAs. The D5 genome encodes genes required for this function, including RNase P subunits P29 (*rpp29*, D5\_0690), P30 (*rpp30*, D5\_1688), P14 (*rpp14*, D5\_1687), and RPR2 (*rpr2*, D5\_2348), ribonuclease Z (*rnz*, D5\_2013), tRNA nucleotidyltransferase Cca (*cca*, D5\_1466), tRNA intron endonuclease EndA (*endA*, D5\_1660), 2'-5' RNA ligase (*ligT*, D5\_1467), tRNA<sup>His</sup> guanylyltransferase ThgL (*thgL*, D5\_1966) and RNA-splicing ligase RtcB (*rtcB*, D5\_0920). The aminoacyl-tRNA synthetases are responsible for linking amino acids to corresponding tRNAs for amino acids and these are listed in Table A.6.1.

*Maturation of rRNA.* The endonuclease and ribonuclease involved in tRNA maturation are also involved in rRNA maturation. In addition, the D5 genome has six genes encoding exosomes subunits (D5\_1685, D5\_1748), RNA-binding protein Rrp4 (D5\_1684), Rrp42 (D5\_1682) and Csl4 (D5\_2432), exonuclease Rrp41 (D5\_1683).

*RNA modification.* D5 is predicted to possess genes involved in RNA post-transcriptional modifications, including tRNA modification to wyosine, quenosine and archaeosine as shown in Table A.6.1. Post-transcriptional modifications, such as methylation and pseudouridylation, are commonly observed on rRNA, tRNA and snRNAs. The ribonucleoprotein (RNP) H/ACA box RNP complex is responsible for pseudouridylation of nucleotides, and three protein subunits have been predicted in the D5 genome, Nop10p (D5\_0169), Cbf5p (D5\_0710) and Gar1p (D5\_1932). The H/ACA box RNP interacts with ribosomal protein L7Ae (D5\_1656). The snoRNA that associates with H/ACA box RNP remains undetected.

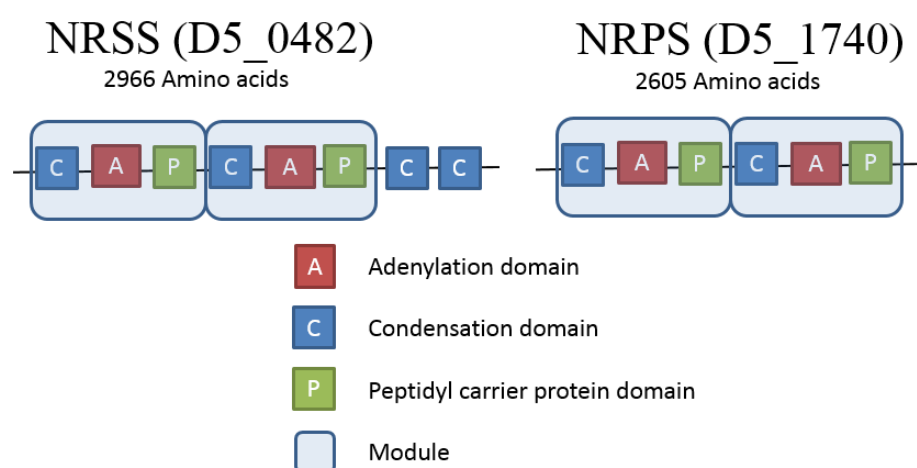
*Ribosomal proteins.* The constituent of the large (50S) subunit and small (30S) subunit of ribosome are listed within Table A.6.1.

*Translation factors.* Translation factors are essential in facilitation of translational initiation, peptide elongation and termination (Table A.6.1). Two modified amino acids have been observed in the translation factors, including deoxyhypusine found in translation initiation factor 5A and diphthamide in elongation factor 2.

*Degradation of mRNA.* The aforementioned exosome complex not only mediates the maturation of rRNA, but it also facilitates in degradation of defective mRNAs. In addition, a NMD3 family protein (D5\_2035) is involved in nonsense-mediated mRNA decay.



*Non-ribosomal peptide biosynthesis systems.* The D5 genome is predicted to encode a large non-ribosomal peptide synthase (NRPS) of 2,605 amino acids (D5\_1740), and an even larger non-ribosomal surfactin synthase (NRSS) of 2,966 aa (D5\_0482). Adjacent to the NRPS gene is a MatE efflux family protein (D5\_1739) that might be involved in export of the NRPS product. Domain analysis of the D5 NRPS reveals two modules, each containing an adenylation domain, a condensation domain and ending with a peptidyl carrier protein domain. The NRSS is predicted with two additional condensation domains near the 3' terminal (Figure 6.23). The substrate specificity for adenylation domains were predicted from a database of adenylation domains with known substrate, and the predicted D5 NRPS adenylation substrate is phenylalanine for both of the adenylation domains. The predicted D5 NRSS adenylation substrates are proline and pipecolic acid.



**Figure 6.23 Non-ribosomal surfactin synthase and non-ribosomal peptide synthase of D5.** The diagram is not drawn to scale.

### *Protein fate*

*Secretion.* The genes predicted to be involved in protein secretion are summarised in Table 6.17. The secretion mechanism utilised by D5 is similar to the other methanogenic archaea, depending on the *N*-terminus signal sequence and the universally conserved Sec pathway. A gene encoding a sortase family protein (D5\_0315) has been predicted in the D5 genome. Sortase acts on secreted proteins with an LPxTG-like sorting motif at the C-terminal and fix the extracellular proteins to the pseudomeurein by cleaving between threonine and glycine, and catalyse amide bond formation between threonine and the cell wall. However, there are no LPxTG/NPxTG/LPxTA/LAxTG motifs predicted amongst the proteins encoded by the D5 genome. Three genes homologous to subunit E and F of type II secretion system have been predicted in the D5 genome.

**Table 6.17 D5 genes predicted to be involved in secretion**

<b>Locus_tag</b>	<b>Predicted product</b>
D5_0315	sortase family protein
<b>Signal recognition particle</b>	
D5_2338	signal recognition particle receptor FtsY
D5_0815	signal recognition particle SRP19 protein
D5_2438	signal recognition particle SRP54 protein
<b>Sec pore</b>	
D5_0704	preprotein translocase subunit SecY
D5_2226	preprotein translocase subunit SecE
D5_2426	preprotein translocase subunit SecG
D5_2360	oligosaccharyl transferase
<b>Type II secretion system</b>	
D5_1979	type II secretion system protein F GspF
D5_2197	type II secretion system protein E GspE
D5_2198	type II secretion system protein F GspF
<b>Signal peptidase</b>	
D5_0046	signal peptidase I
D5_1565	type IV leader peptidase family protein
D5_1673	signal peptidase I

*Protein folding and degradation.* The D5 genome is predicted to encode a variety of chaperones and peptidases (Table A.6.1). The genes involved in assisting protein folding and denaturing of misfolded proteins are similar to what was described for ISO4-H5 in Chapter 3, Section 3.2.4.

### *Regulation*

The D5 genome is predicted to encode 21 transcriptional regulators, most have unknown functions (Table A.6.1).

### *Oxidative stress*

*Methanobrevibacter* sp. D5 is an obligate anaerobic methanogen, and some of its enzymes are highly sensitive to oxygen, such as the core enzyme complex of methanogenesis Mcr. Therefore *Methanobrevibacter* sp. D5 must carry genes allowing it to combat oxidative stress. The D5 genome is predicted to possess a full complement of genes allowing D5 to synthesise and utilise glutathione from glutamate, cysteine and glycine (Table 6.18). Glutathione peroxidase detoxifies H<sub>2</sub>O<sub>2</sub> using reduced glutathione as an electron donor. Glutathione also restores oxidised molecules by acting as a hydrogen donor. No superoxide dismutases were predicted in the D5 genome. Oxygen can be detoxified via F<sub>420</sub>H<sub>2</sub> oxidase, FprA, to H<sub>2</sub>O or detoxified to H<sub>2</sub>O<sub>2</sub> by desulfoferrodoxin or NADH oxidase. The H<sub>2</sub>O<sub>2</sub> is detoxified by rubrerythrin to H<sub>2</sub>O.

**Table 6.18 D5 genes involved in oxidative stress**

<b>Locus_tag</b>	<b>Predicted product</b>
<b>O<sub>2</sub> detoxification</b>	
D5_1478	F <sub>420</sub> H <sub>2</sub> oxidase FprA
D5_2452	F <sub>420</sub> H <sub>2</sub> oxidase FprA
D5_0115	NADH oxidase Nox
D5_1416	NADH oxidase Nox
D5_1696	desulfoferrodoxin Dfx
<b>H<sub>2</sub>O<sub>2</sub> detoxification</b>	
D5_1600	rubredoxin Rub1
D5_1601	rubredoxin Rub2
D5_1622	rubrerythrin Rbr1
D5_1623	rubrerythrin Rbr2
D5_1624	rubrerythrin Rbr3
D5_2257	rubredoxin Rub3
D5_2453	rubrerythrin Rbr4
<b>Thioredoxin</b>	
D5_0599	thioredoxin TrxA
D5_1807	thioredoxin-disulfide reductase TrxB
<b>Glutathione</b>	
D5_0781	gamma-glutamylcysteine synthetase GshA
D5_1281	glutathione peroxidase GpxA
D5_1347	glutathione-disulfide reductase Gor
D5_1806	glutathione peroxidase GpxA
D5_2236	glutathione-disulfide reductase Gor
D5_2270	bifunctional glutamate-cysteine ligase/glutathione synthetase GshF

### 6.2.6 Comparative analysis of methanogenesis pathway in *Methanobrevibacter* spp.

The complete comparison of all metabolic pathways between 18 *Methanobrevibacter* genomes was not possible during the timeframe of this thesis, but a detailed analysis of the hydrogenotrophic methanogenesis pathway was performed. Hydrogenotrophic methanogenesis is the sole energy generating mechanism of the *Methanobrevibacter* spp. analysed. The genes predicted to be involved in methanogenesis are summarized in Figure 6.21 and Table A.6.9.

The observations from culture based experiments has shown all *Methanobrevibacter* spp. utilise CO<sub>2</sub> and H<sub>2</sub> as substrates for methanogenesis (Table 6.1), this is reflected by the full complement of genes required to carry out hydrogenotrophic methanogenesis in all of the genomes analysed. The H<sub>2</sub>-dependent fixation of CO<sub>2</sub> is carried out by a formylmethanofuran dehydrogenase Fwd, which binds CO<sub>2</sub> to a methanofuran (MF) and reduces it to a formyl-MF. There are two isozymes of formylmethanofuran dehydrogenase encoded by two gene clusters, the *fwdHFGDACB* operon that is transcribed in the presence of tungstate or molybdate and a *fmdECB* operon that is transcribed in the presence of molybdate. Only the *fwd* operon was identified in the *Methanobrevibacter* spp. genomes. The formyl group on formyl-MF is then

transferred to the tetrahydromethanopterin (H<sub>4</sub>MPT) by formylmethanofuran-tetrahydromethanopterin formyl transferase Ftr, producing formyl-H<sub>4</sub>MPT. The formyl-H<sub>4</sub>MPT is reversibly catalysed to N<sup>5</sup>,N<sup>10</sup>-methenyl-H<sub>4</sub>MPT by N<sup>5</sup>,N<sup>10</sup>-methenyl-H<sub>4</sub>MPT cyclohydrolase Mch. The JMR01 genome is predicted to possess an additional copy of the *mch* gene. Methenyl-H<sub>4</sub>MPT is reduced to methylene-H<sub>4</sub>MPT by either F<sub>420</sub>-dependent methylene-H<sub>4</sub>MPT dehydrogenase, Mtd or H<sub>2</sub>-dependent methylene-H<sub>4</sub>MPT dehydrogenase, Hmd. All of the *Methanobrevibacter* genomes were predicted to possess the *mtd* gene, with an additional copy of the *mtd* gene predicted in the D5, ZA-10<sup>T</sup> and SM9 genomes. The *hmd* gene was not found in the RFM-1<sup>T</sup>, RFM-2<sup>T</sup>, RFM-3<sup>T</sup>, ATM<sup>T</sup> and ANOR1 genomes, whereas YE315, ZA-10<sup>T</sup> and SM9 genomes are predicted to possess an additional copy of the *hmd* gene. Methylene-H<sub>4</sub>MPT is subsequently reduced to methyl-H<sub>4</sub>MPT by Methylene-H<sub>4</sub>MPT reductase, Mer. Both Mer and Mtd acquire reducing potential from the F<sub>420</sub> hydrogenase, Frh, encoded by *frhADGB*, and all *Methanobrevibacter* genomes have the Frh-encoding genes. The methyl group is transferred from methyl-H<sub>4</sub>MPT to CoM via tetrahydromethanopterin S-methyltransferase complex, Mtr, which is coupled to the generation of Na<sup>+</sup> gradient across the cytoplasmic membrane. The methyl-CoM is reduced to CH<sub>4</sub> by methyl-CoM reductase, Mcr/Mrt, all genomes were predicted to possess genes encoding Mcr, whereas the genes encoding Mrt was only predicted in the D5, ZA-10<sup>T</sup>, SM9, HO<sup>T</sup>, PS<sup>T</sup>, JMR01, KM1H5-1P<sup>T</sup> and YLM1 genomes.



involved in methanogenesis except the M1<sup>T</sup> genome (Figure 6.24, Table A.6.9), which lacks the *comADE* genes required for *de novo* CoM biosynthesis.

The reducing potential required for methanogenesis is supplied by Frh and F<sub>420</sub>H<sub>2</sub> independently of membrane potential, all genomes analysed are predicted to possess a full set of genes encoding hydrogenases Frh, MvH and EhaA involved in supplying reducing potential for methanogenesis, and possess a full set of genes encoding A<sub>1</sub>A<sub>0</sub> ATP synthase that utilise the membrane potential for energy generation. In addition, a gene encoding NADP-dependent F<sub>420</sub> reductase, NpdG (D5\_1461) has been predicted in the all *Methanobrevibacter* genomes, however, the NADP dependent alcohol dehydrogenase, *adh* gene, that allows *Methanobrevibacter* spp. to utilise alcohol as reducing potential, is only predicted in the YE315, M1<sup>T</sup>, KM1H5-1P<sup>T</sup> and YLM1 genomes.

### 6.3 Discussion

There are 15 type strains of *Methanobrevibacter* spp, all utilise CO<sub>2</sub> and H<sub>2</sub> as a substrate for hydrogenotrophic methanogenesis. In addition, *Mbb. smithii* PS<sup>T</sup>, *Mbb. ruminantium* M1<sup>T</sup>, *Mbb. woesei* GS<sup>T</sup>, *Mbb. olleyae* KM1H5-1P<sup>T</sup>, *Mbb. millerae* ZA-10<sup>T</sup>, *Mbb. boviskoreani* JH1<sup>T</sup>, *Mbb. arboriphilus* DH1<sup>T</sup> and *Mbb. cuticularis* RFM-1<sup>T</sup> are capable of using formate. Other strain dependent nutrient requirements such as CoM and acetate have been detailed in Table 6.01. In culture studies with D5, a slightly higher growth has been observed when acetate (20 mM) was supplemented in the media. *Methanobrevibacter* sp. D5 is reported to have a H<sub>2</sub> threshold similar to *Mbb. ruminantium* M1<sup>T</sup>, *Methanobrevibacter* sp. AbM4 and *Mbb. millerae* SM9 (Kim 2012), which suggests the growth conditions and niche occupied by these strains are highly similar.

This chapter describes the sequencing, gap closing and bioinformatics analysis of the D5 genome. D5 is a member of the *Mbb. gottschalkii* clade isolated from the ovine rumen (Seedorf *et al.* 2014), and its genome sequence represents an important contribution to the knowledge accumulating around this dominant group of rumen methanogens. Its 2.8 Mb genome has provided new insight into the lifestyle and cellular processes of this organism. At the beginning of this project, no genome sequence was available for methanogens of the *Mbb. gottschalkii* clade. Currently (July 2016), four other members of the *Mbb. gottschalkii* clade have been sequenced, and 17 genome sequences of *Methanobrevibacter* spp. in total are included in this study. This chapter also compares the D5 genome to the other available *Methanobrevibacter*

spp. genomes and has identified the core genome that is important to the survival of *Methanobrevibacter* spp.

The assembly confirmation indicated a discrepancy predicted to be caused by an undigested *ApaI* site at 226,045 bp (Figure 6.5). Base 226,045 is within scaffold one, contig 26, and was not within a gap region, therefore the *ApaI* site was not introduced by PCR sequencing error. The *ApaI* restriction endonuclease cleavage may be inhibited by DNA-cytosine methyltransferase (Dcm) catalysed methylation at CpG (Larimer 1987). D5 is predicted to possess two *dcm* genes (D5\_1219, D5\_1220), therefore it is plausible that the *ApaI* restriction recognition site at 226,045 bp may have been methylated and prevented the *ApaI* cleavage. The 5-methylcytosine can be spontaneously deaminated to thymidine (Ehrlich *et al.* 1986), the resulting TG mismatch base pairs can be repaired by a short patch endonuclease Vsr (D5\_1221) (Lieb 1991), which is encoded adjacent to the *dcm* genes.

A 170 kb band was observed in the undigested lane of D5 genomic DNA (Figure 6.5). This result would typically indicate an extrachromosomal element, however, no contig of 170 kb in size was observed within the D5 assembled genome. Within the assembled sequence derived from the Illumina sequencing data, a second genome was present. The contaminant was identified as a member of the genus *Bacillus*, therefore it is possible that the extrachromosomal element belongs to the *Bacillus* contaminant, as no *Methanobrevibacter* spp. have been reported of carrying extrachromosomal elements thus far (Samuel *et al.* 2007; Leahy *et al.* 2010; Leahy *et al.* 2013; Lee *et al.* 2013; Khelaifia *et al.* 2014; Khelaifia *et al.* 2014; Kelly *et al.* 2016).

Within the assembled sequence derived from the Illumina data, a contig of 10,943 bp was found to have a read coverage four times that of the D5 contigs and over 300 times higher read number than the contaminating *Bacillus* contigs (Figure 6.3). This result is often suggestive of an extrachromosomal element. The difference in read number suggested that this contig may be more likely to be associated with the D5 rather than the contaminating *Bacillus*, although some plasmids such as pUC19 from *Escherichia coli* may have as high as 700 copies per cell (Casali and Preston 2003). This contig has a %G+C of 27.6%, which is more similar to D5 (33.1%) than the *Bacillus* contaminant (37.4%). Attempts to circularise the contig via PCR have been unsuccessful. Furthermore, this contig was identified within the enrichment culture H6 sequenced by 454 pyrosequencing (Figure 6.2). PCR was able to identity the persistence of this contig within the D5 culture (result not shown). This contig encodes 16 genes; 14 are

hypothetical proteins with no homology to known genes, and two genes are homologous to an rRNA methyltransferase, *rlmN* gene (Tsac\_1724) and a stage II sporulation protein E gene (Tsac\_1725) of *Thermoanaerobacterium saccharolyticum* JW/SL-YS485 (GenBank: CP003184.1). *T. saccharolyticum* harbors an active prophage, THSA-485A, but the prophage region does not encompass the two homologous genes (Currie *et al.* 2015). Prophage and transposable elements have terminal repeats that facilitate their integration and excision by recombination (Levis *et al.* 1980; Temin 1981; Dejean *et al.* 1984). A direct repeat of 127 bp (ATGGGTTTATCAAATGATATTAAATTTTCAATCAATGTCTTATGGAATAATACTT TAATATATAGAAAACATTTAATTAATTTAAATGACTTAATACAAAGATTTAGAGT TTATAATAAAAATCAG) with 100% nucleotide sequence identity is present at base 1 and base 8975 of the contig. This repeat sequence displayed a 68% pairwise identity to a region upstream to a transposase (D5\_1474) in the D5 genome, and 64.6% pairwise identity to a region on Node\_163 of the contaminating *Bacillus* contigs. The true identity and association of this contig remains unknown. The presence of prophage has been shown in four *Methanobrevibacter* genomes, including M1<sup>T</sup>, SM9, YLM1 and PS<sup>T</sup>, while no *Methanobrevibacter* genomes have plasmids. Therefore it is more likely that this contig may belong to a stray prophage rather than a plasmid sequence.

Eighteen genomes of *Methanobrevibacter* spp. sourced from four different gut environments (human, bovine, ovine and termite) and from an anaerobic digester have been compared in this study, to identify the extent of gene conservation and to gain insight into the metabolism and physiology of this group of methanogens. Amongst the sequenced *Methanobrevibacter* spp., the genome size ranges between 1.66 Mbp (ATM<sup>T</sup>) and 2.94 Mbp (M1<sup>T</sup>), and the %G+C range is between 24.2% (SH<sup>T</sup>) and 36.4% (ZA-10<sup>T</sup>). The genomes all contain circular chromosomes with no extrachromosomal elements. Seven of the genomes are complete, including D5, YE315, SM9, M1<sup>T</sup>, PS<sup>T</sup>, YLM1 and AbM4.

The phylogenetic relationship of members of the *Methanobrevibacter* can be inferred from their 16S rRNA gene sequences (Seedorf *et al.* 2014) (Figure A.6.1). The inferred relationships suggests that the organisms D5, YE315, ZA-10<sup>T</sup>, SM9 and HO<sup>T</sup> are members of the *Mbb. gottschalkii* clade, PS<sup>T</sup> is in the *Mbb. smithii* clade, JMR01 is in the *Mbb. oralis* clade, M1<sup>T</sup>, KM1H5-1P<sup>T</sup> and YLM1 are members of the *Mbb. ruminantium* clade, SH<sup>T</sup>, AbM4 and JH1<sup>T</sup> are members of the *Mbb. wolinii* clade, while ANOR1, RFM-1<sup>T</sup>, RFM-2<sup>T</sup>, RFM-3<sup>T</sup> and ATM<sup>T</sup> are placed in the *Mbb. arboriphilus* clade, *Mbb. cuticularis* clade, *Mbb. curvatus* clade, *Mbb. filiformis* clade and *Mbb. acididurans* clade, respectively. RFM-1<sup>T</sup>, RFM-2<sup>T</sup>, RFM-3<sup>T</sup> and



ANOR1 are the most distantly related to other *Methanobrevibacter* spp. (Figure 6.6). Recent advancements in bioinformatics methods recommends that a 16S rRNA gene identity of  $\leq 98.7\%$  can be considered as a separate species (Stackebrandt and Ebers 2006; Petti 2007; Schlager *et al.* 2012). Using this recommendation, *Mbb. boviskoreani* JH1<sup>T</sup> and *Methanobrevibacter* sp. AbM4 can be considered as different strains of the same species, *Methanobrevibacter* sp. D5, *Mbb. millerae* ZA-10<sup>T</sup> and *Methanobrevibacter* sp. YE315 can be considered as different strains of the same species, *Mbb. olleyae* KM1H5-1P<sup>T</sup> and *Mbb. olleyae* YLM1 can be considered as different strains of the same species (Table 6.4). A FGD analysis and a pangenome tree reveals a different profile (Figure 6.7 and 6.8) where JMR01 and PS<sup>T</sup> are clustered together with members of the *Mbb. gottschalkii* clade. This suggests JMR01 and PS<sup>T</sup> are similar to members of the *Mbb. gottschalkii* clade at the ORFeome level, whereas for members of the *Mbb. ruminantium* clade and *Mbb. wolinii* clade, their ORFeome similarity is consistent with the inferred phylogenetic relationship.

A reported feature of genomes from the *Methanobrevibacter* spp. is the organisation of rRNA genes into an operon (Samuel *et al.* 2007; Leahy *et al.* 2010), which is typically ordered 16S, 23S and 5S rRNA genes (Brosius *et al.* 1981). Archaeal rRNA operons vary in both structure and frequency (Dennis 1997), with 50% of sequenced archaeal genomes containing only one rRNA operon, but with some having as many as four operons (Yip *et al.* 2013). The rRNA operons in the *Methanobrevibacter* genomes analysed in this study contain between one to three rRNA operons. Some rRNA operons may contain one or two tRNAs within spacer sequences, including tRNA<sup>Glu</sup>, tRNA<sup>Ile</sup>, tRNA<sup>Ala</sup> and tRNA<sup>Cys</sup> (Lund *et al.* 1976; Ikemura and Nomura 1977; Morgan *et al.* 1977; Keller *et al.* 1980; Chant and Dennis 1986). Some archaeal genomes are found with the 5S rRNA gene existing outside the rRNA operon, as observed in the *Desulfurococcus mobilis*, *Sulfolobus solfataricus* and *Methanococcus jannaschii* genomes (Kjems and Garrett 1987; Bult *et al.* 1996; She *et al.* 2001). A tRNA<sup>Ala</sup> has been identified in one of the rRNA operons in the D5 genome, between the 16S rRNA gene and the 23S rRNA gene. The arrangement of tRNA<sup>Ala</sup> between the 16S and the 23S rRNA genes have also been identified in other completed *Methanobrevibacter* genomes. YE315, SM9 and PS<sup>T</sup> only have one tRNA<sup>Ala</sup> containing rRNA operon, the tRNA<sup>Ala</sup> is present in both rRNA operons of M1<sup>T</sup> and YLM1 genome, and in the AbM4 genome the tRNA<sup>Ala</sup> is found in two of the three rRNA operons. The arrangement of tRNA<sup>Ala</sup> in the rRNA operon may reflect the phylogenetic relationship of the *Methanobrevibacter* spp., as both YLM1 and M1<sup>T</sup> are members of the *Mbb.*

*ruminantium* clade, and D5, YE315 and SM9 are members of the *Mbb. gottschalkii* clade (Seedorf *et al.* 2014).

The genome of D5 encodes 44 tRNAs encoding all amino acids. All except one of the *Methanobrevibacter* genomes analysed are predicted to encode a full set of tRNAs corresponding to 20 amino acids. The KM1H5-1P<sup>T</sup> genome is predicted to lack a tRNA<sup>Ala</sup>, however the KM1H5-1P<sup>T</sup> genome is incomplete and thus the tRNA<sup>Ala</sup> may be contained within the missing region. Introns in tRNAs are found in all three kingdoms of life and are present on average in 15% of archaeal tRNAs (Yoshihisa 2014). Two of the D5 tRNAs, tRNA<sup>Met</sup> and tRNA<sup>Trp</sup>, contain an intron. The tRNA intron endonuclease *endA* (D5\_1660) required for tRNA intron removal (Thompson *et al.* 1989) has been identified within the D5 genome. All of the *Methanobrevibacter* genomes analysed are predicted to encode intron containing tRNA<sup>Met</sup> and tRNA<sup>Trp</sup>, which suggests this may be a conserved trait to the *Methanobrevibacter* spp. Similar to D5, all the genomes analysed are predicted to encode an *endA* gene. Sixteen of the genomes analysed possess similar number of tRNAs, ranging from 33 (YE315, JMR01, KM1H5-1P<sup>T</sup>, YLM1, RFM-3<sup>T</sup> and ATM<sup>T</sup>) to 44 (D5). There are two genomes predicted with exceptionally high number of tRNAs, M1<sup>T</sup> with 58 and ZA-10<sup>T</sup> with 77, attributed partially by a high copy number of pseudo tRNAs and tRNA<sup>Leu</sup>. M1<sup>T</sup> and ZA-10<sup>T</sup> also possess an additional intron-containing tRNA, a tRNA<sup>Lys</sup> in M1<sup>T</sup> and a tRNA<sup>Ser</sup> in ZA-10<sup>T</sup>.

The codon and amino acid usage of the genomes was compared and found to be similar between the 18 *Methanobrevibacter* spp. (Figure 6.11, Table A.6.2). A higher usage of asparagine was observed in the termite sourced RFM-1<sup>T</sup>, RFM-2<sup>T</sup> and RFM-3<sup>T</sup>. M1<sup>T</sup> and ZA-10<sup>T</sup> displayed a mildly different codon usage pattern, and incidentally, these two genomes also possess an unusually high number of tRNAs with many clustered together. It was found that laterally transferred genes may carry codons not used by the host, and thereby prompting the selective retention of foreign tRNAs in the host genome (McDonald *et al.* 2015). M1<sup>T</sup> was found to possess genes that are likely to have been horizontally acquired, including a NRPS (Leahy *et al.* 2010). Therefore it is plausible that ZA-10<sup>T</sup> may also possess horizontally transferred genes that has contributed to the difference in codon usage and the high copy number of tRNAs.

The CRISPR/Cas system facilitates acquired immunity (Marraffini and Sontheimer 2010) and has been classified into three major types and ten subtypes based on the signature genes present (Makarova *et al.* 2011). Both Type I and Type II contain the central *cas1* and *cas2* genes required for spacer sequence insertion, while the signature gene for Type I CRISPR/Cas system

is a *cas3* gene that cleaves the invading DNA. For Type II the signature gene is the large *cas9* gene that utilises trans-encoded small RNA (tracrRNA) to guide the processing of pre-crRNA. The signature gene for Type III is *cas10* gene, and many Type III were found to lack *cas1* and *cas2* genes (Makarova *et al.* 2011; Makarova *et al.* 2011). While most of the CRISPR/Cas system can be readily classified into the three major types, some CRISPR/Cas system lacks signature genes to all three major types, and remain unclassified as Type U (Makarova *et al.* 2011). The CRISPR/Cas system of *Methanobrevibacter* sp. D5 lacks the signature genes to all three major types, and is therefore classified as type U. The presence of *cas10* and *csm3* genes in SM9, JMR01, M1<sup>T</sup>, KM1H5-1P<sup>T</sup>, YLM1, SH<sup>T</sup>, RFM-2<sup>T</sup> and RFM-3<sup>T</sup> suggested these genomes possess a type III-A CRISPR system. The presence of *cas3* gene in YE315, ZA-10<sup>T</sup>, PS<sup>T</sup>, AbM4, JH1, RFM-1<sup>T</sup> and ATM<sup>T</sup> suggested these genomes possess a type I CRISPR system, and the genes present are insufficient to further classify into subclasses. The *cas3* and *cas10* genes were not identified in the HO<sup>T</sup> genome, which classifies the CRISPR elements in the HO<sup>T</sup> genome as type U. Interestingly, despite the presence of CRISPR associated genes in the ZA-10<sup>T</sup> and HO<sup>T</sup> genomes, no CRISPR elements were identified in the ZA-10<sup>T</sup> and HO<sup>T</sup> genomes. Since neither genomes were complete, and repeat elements such as CRISPR can often be difficult to sequence, it is possible the CRISPR elements are within the gap regions. No CRISPR associated genes were detected in the ANOR1 genome, which questions whether the ANOR1 genome CRISPR elements identified were genuine. The three predicted CRISPR elements in the ANOR1 genome consist of small number of spacers ranging from two to five, which suggest the identified elements may be other types of repeat elements.

The spacer sequences of the identified CRISPR element were used to identify foreign genetic elements that *Methanobrevibacter* sp. D5 may have encountered in the past (Biswas *et al.* 2013; Shariat *et al.* 2015). Out of 20 spacer sequences analysed, 11 matched known phage and plasmid sequences. Plasmid hosts included *Providencia rettgeri*, *Bacillus cereus* and *Clostridium perfringens*, which can be found in the soil (Yamagishi *et al.* 1964; Kitts *et al.* 1994; Arnesen *et al.* 2008), as well as the parasitic Periwinkle leaf yellowing phytoplasma found in plants (Nejat *et al.* 2013). It is conceivable that within the rumen environment *Methanobrevibacter* sp. D5 may have come into contact with similar organisms, as ruminants ingest soil while grazing. Furthermore, phytoplasma may have colonised the forage plant material, and exposed the grazing ruminant to their mobile genetic elements.

Horizontal gene transfer occurs commonly amongst prokaryotes, and continuous acquisition and removal of genes in the genome is an integral part of genome evolution (Ochman *et al.*

2000; Koonin *et al.* 2001), assisting with environmental adaptation (Gilbert and Cordaux 2013). The genes acquired from horizontal transfer often differ to other genes in the genome in base composition, codon usage and bias of di- and tri-nucleotide frequencies (Ochman *et al.* 2000). The regions that may contain horizontally transferred DNA and putatively horizontally transferred genes have been predicted in the D5 genome. However, most of the regions have a low IVOM score as well as a high average CAI which suggests the likelihood of containing horizontally transferred DNA is low. This is consistent with the prediction of putative horizontally transferred genes, as only 15 regions contain predicted putative horizontally transferred genes. Furthermore, some regions contain genes predicted to be involved in important metabolism; the two regions with highest IVOM score is actually where the rRNA operons are found, and they are unlikely to be horizontally transferred. Region 30 had the third highest IVOM score with a predicted transposase within and two hypothetical genes that are also predicted to be horizontally transferred.

IS elements are segments of DNA, less than 2.5 Kb in length, and capable of inserting into a target sequence (Frost *et al.* 2005). IS elements often encode a transposase flanked with short inverted repeats necessary for transposase recognition and cleavage (Mahillon and Chandler 1998). A region of DNA flanked by two IS elements may be mobilized as a composite transposon (Siguier *et al.* 2015). IS elements constitute 1.8% of the D5 genome, whereas IS elements were found to constitute only 0.2% of the M1<sup>T</sup> genome (Leahy *et al.* 2010). Higher number of IS elements may act as a driving force of genome evolution (Biemont and Vieira 2006), aiding D5 to persist and define its niche within the competitive habitat of the rumen environment.

The *Methanobrevibacter* cells are resistant to lysis by SDS (Miller and Lin 2002), and to various antibiotics that targets bacterial cell wall, including but not limited to vancomycin, penicillin and cephalothin (Dermoumi and Ansorg 2001). These properties are conveyed by possession of pseudomurein in their cell wall and the triple-layered structure of their cell envelopes (Zeikus and Henning 1975; Kandler and König 1978), with an electron dense, thin pseudomurein layer outside the cell membrane, followed by a thicker but less tightly packed pseudomurein layer surrounding it, and a cell wall glycopolymer at outermost layer (Leahy *et al.* 2010). The specific genes involved in pseudomurein biosynthesis has been proposed but still require experimental characterisation (Kandler and König 1998). In addition to pseudomurein, the sugar content of the cell wall includes galactose, rhamnose and low levels of glucose, mannose (Kandler and König 1978), which may constitute the glycopolymer outer

layer. Several genes may be involved in the biosynthesis of teichoic acid, sialic acid and exopolysaccharide, as listed in Table A.6.1.

Extracellular and surface-associated proteins are likely to play an important role in many essential interactions and adaptations to the environment. D5 is predicted to use the SRP, Sec61 translocation system for translocation of gene products (Zwieb and Bhuiyan 2010). Although D5 is predicted to encode a sortase family protein (D5\_0315), no LPxTG/NPxTG/LpxTA/LAxTG sorting motif harboring genes were detected (Ton-That *et al.* 2004), which suggests the sortase family protein may not be a functional sortase. Three homologs (D5\_1979, D5\_2197, D5\_2198) of a type II secretion system were predicted in the D5 genome, however, a functional type II secretion system typically involves at least 12 different gene products (Sandkvist 2001), which suggests D5 does not carry a functional type II secretion system. The three genes homologous to the type II secretion system are conserved across the *Methanobrevibacter* genomes, with an average amino acid identity of 59.1%, 68.6% and 52.2% respectively for the corresponding D5\_1979, D5\_2197 and D5\_2198 genes. The twin-arginine translocation system only requires two subunits to export folded gene products (Ellen *et al.* 2010), however no homologous genes were identified in the D5 genome. Another type of secretion mechanism employed by archaea is the formation of a vesicle by ESCRT-III (endosomal sorting complex required for transport) proteins (Ellen *et al.* 2009), but no homologs of ESCRT gene were predicted in the D5 genome. There are 173 D5 genes predicted to encode an N-terminal signal peptide (Figure 6.12), 33 proteins are predicted with two or more transmembrane domain, and an additional 103 proteins are predicted to be anchored through a N-terminal transmembrane domain, two proteins are predicted to be anchored through a C-terminal transmembrane domain. In the genomes analysed, an average of 4.9% of their ORFeomes are devoted to the process of protein export. D5 is predicted to have the highest proportion of exported gene products, 7% of its ORFeome, while SH<sup>T</sup> is predicted to have the lowest proportion of exported gene products, 3.5% of its ORFeome (Table 6.12). Proteins predicted to contain a signal peptide without a transmembrane helice or lipobox are thought to be secreted into the environment. The number of proteins predicted to be secreted into the environment range between 3 and 26 per genome for the *Methanobrevibacter* genomes analysed. Extracellular proteins with domain repeats are often involved in binding to the cell (Cabanes *et al.* 2002), and all of the genomes examined contain genes with predicted repeat domains (27, 19, 22, 26, 5, 11, 7, 37, 9, 10, 3, 2, 3, 12, 3, 9, 2 and 6 genes for D5, YE315, ZA-10<sup>T</sup>, SM9, HO<sup>T</sup>, PS<sup>T</sup>, JMR01, M1<sup>T</sup>, KM1H5-1P<sup>T</sup>, YLM1, SH<sup>T</sup>, AbM4, JH1<sup>T</sup>, RFM-1<sup>T</sup>, RFM-

2<sup>T</sup>, RFM-3<sup>T</sup>, ATM<sup>T</sup> and ANOR1 respectively). Amongst the secretome, there are four other Pfam domains conserved across all the genomes analysed, these domains are PF09972 DUF2207, PF13229 Right handed beta helix, PF02516 Oligosaccharyl transferase STT3 and PF04206 Tetrahydromethanopterin S-methyltransferase subunit E (MtrE) (Table A.6.04). The function of DUF2207 is unknown, right handed beta helix is a solenoid protein domain formed by parallel beta strands in helical arrangement, its structure has been found within pectate lyase (Jenkins *et al.* 2004), surface adhesin PfbA of *Streptococcus pneumoniae* (Beulin *et al.* 2014) and periplasmic alginate epimerase AlgG (Wolfram *et al.* 2014). The oligosaccharyl transferase STT3 domain has been known to function in *N*-glycosylation of the S-layer glycoprotein in *Haloferax volcanii* (Abu-Qarn *et al.* 2007), this suggests all *Methanobrevibacter* spp. utilise a *N*-glycosylated pseudomurein. The Mtr is essential to hydrogenotrophic methanogenesis and contributes to generation of the Na<sup>+</sup> gradient, as all of the *Methanobrevibacter* spp. are hydrogenotrophic methanogens, it is conceivable for all genomes analysed to encode a transmembrane protein with signal peptide predicted with the MtrE domain.

Adhesins are large, repeat-rich cell surface proteins that mediate adhesion to other cells or surfaces and biofilm formation (Fuqua 2010). Adhesins are responsible for methanogen adherence to hydrogen producers such as protozoa (Finlay *et al.* 1994). Adhesin-like proteins have been identified in *Mbb. smithii* and *Mbb. ruminantium* M1<sup>T</sup>, and *Mbb. smithii* has adhesin-like proteins homologous to pectin esterase and mediates in binding of human mucosal chondroitin (Samuel *et al.* 2007). *Mbb. ruminantium* M1<sup>T</sup> was found to co-aggregate with *Butyrivibrio proteoclasticus* in coculture, and six adhesin-like proteins were upregulated (Leahy *et al.* 2010). One of the upregulated adhesin-like proteins (mru\_1499) was found to be a broad spectrum adhesin important in binding of M1<sup>T</sup> to a broad range of rumen protozoa (Ng *et al.* 2016). D5 is predicted to encode 82 adhesin-like proteins, which is less than M1<sup>T</sup> (105), SM9 (95) but more than PS<sup>T</sup> (48) and AbM4 (29). The mru\_1499 adhesin-like protein is predicted to harbor a N-terminal signal peptide, four Big\_1 (PF02369: bacterial Ig-like (group 1)) domains and a C-terminal transglutaminase-like domain (PF01841) (Ng *et al.* 2016). Using phage display, the N-terminal Big\_1 domain was discovered to have an important role in cell binding (Ng *et al.* 2016). The adhesin-like proteins predicted with transglutaminase domain was proposed to mediate cell adherence by cross-linking (Griffin *et al.* 2002), which was later revealed to be less important than big\_1 domains in cell adherence (Ng *et al.* 2016), and has been proposed that it may be involved in modification of the cell surface (Leahy *et al.* 2010). Of the 82 predicted adhesin-like proteins in D5, 16 are predicted to harbor Big\_1 domains, of

which three are also predicted to harbor a transglutaminase-like domain, including a D5\_1662 gene that is homologous to mru\_1499. Therefore, it is conceivable that D5 may possess a similar cell binding ability to M1<sup>T</sup>, capable of binding to H<sub>2</sub> producing protozoa and bacteria in the rumen. In addition to D5, adhesin-like proteins homologous to mru\_1499 has been identified in KM1H5-1P<sup>T</sup>, YLM1, PS<sup>T</sup>, ZA-10<sup>T</sup>, SM9, AbM4 and *Methanosphaera stadmanae* MCB-3 (Ng *et al.* 2016), which suggests conserved domains. The adhesins maybe a crucial component of how the *Methanobrevibacter* spp. interact with their environment, the difference in the microorganism or host gut it binds to, assists them in establishing their own niche in the presence of closely related species.

The conservation in genome structure and gene order between different genomes is known as synteny (Passarge *et al.* 1999). When two genomes have a high degree of synteny, a linear genome alignment can be observed across the entire length of the genome, and when an X-alignment is observed, as reported for *Mycobacterium leprae* and *Mybacterium tuberculosis*, this indicates that matching sequences occur at the same distance from the origin but not necessarily on the same side of the origin (Eisen *et al.* 2000). Using the D5 genome as a reference, it was shown that a high degree of synteny is shared with SM9 and YE315 genomes. An inversion was observed in the YE315 analysis, this inversion encompasses approximately 1 Mb of the genome. A similar inversion was also observed in the SM9 genome that encompasses approximately 2.3 Mb of the genome. The D5 genome has a moderate degree of conservation to M1, AbM4 and YLM1 genome in regions approximately 0.9 Mb from the origin of replication *cdc6-1* gene. However the middle section of the genome lacks synteny, which suggest this region of the genome may harbor clade-specific genes, and the conserved genes between *Methanobrevibacter* spp. tend to lie closer to the origin of replication.

Comparison between *Methanobrevibacter* genomes can reveal conserved genes important for the organisms' growth and survival as well as those genes that might be important for niche adaptation. The core genome of the 18 *Methanobrevibacter* spp. examined consists of 764 gene families, while the order level pan-genome encompasses 9,382 gene families (Figure 6.14). The sequenced *Methanobrevibacter* spp. analysed in this study originated from six environments, ovine (7) and bovine (4) rumen, termite gut (3), anaerobic digester (1), human faeces (2) and subgingival plaque (1). There are two gene families conserved between the rumen strains, which might shed light on the genes involved in adaptation to the rumen environment. Each genome also contains its own unique gene families, from 31 gene families in AbM4 to 649 gene families in JMR01 (Figure 6.16). These unique gene families can offer

insight into how these members of the *Methanobrevibacter* may establish their own niche in the presence of similar organisms within the same environment.

The core genome of *Methanobrevibacter* spp. was classified according to the COG database (Table 6.14). A total of 85 gene families are implicated in energy conservation, which includes the genes required for hydrogenotrophic methanogenesis, and the *fdhBCD* genes involved in formate utilisation, but not the *fdhA* gene, as the *fdhA* gene exists as a pseudogene in the YE315 genome, although it is conserved in 17 other *Methanobrevibacter* spp. genomes analysed. This suggests CO<sub>2</sub>, H<sub>2</sub> and formate are the conserved substrates of *Methanobrevibacter* spp. While experimental findings support the ubiquitous usage of CO<sub>2</sub> and H<sub>2</sub> amongst known *Methanobrevibacter* spp., experimental evidence has shown some members of the genus *Methanobrevibacter* are incapable of utilising formate, including *Mbb. wolinii* SH<sup>T</sup>, *Mbb. gottschalkii* HO<sup>T</sup>, *Mbb. oralis* ZR<sup>T</sup>, *Mbb. curvatus* RFM-2<sup>T</sup>, *Mbb. filiformis* RFM-3<sup>T</sup> and *Mbb. acididurans* ATM<sup>T</sup> (Smith and Hungate 1958; Bryant *et al.* 1971; Balch *et al.* 1979; Ferrari *et al.* 1994; Leadbetter and Breznak 1996; Leadbetter *et al.* 1998; Miller 2001; Miller and Lin 2002; Savant *et al.* 2002; Rea *et al.* 2007; Lee *et al.* 2013). A similar conundrum has been encountered in the *Methanococcus* spp., where the presence of formate utilising genes in the genomes does not correspond to actual utilisation of formate in the cultures (Wood *et al.* 2003). Two sets of formate utilising genes has been found in the *Methanobrevibacter* spp. genomes analysed, one set is the *fdhDCAB* found in all *Methanobrevibacter* spp. genomes analysed, the other is the *flpABD* that was found in 11 of the 18 genomes analysed. The *flpBD* genes are present at end of the contig in the HO<sup>T</sup> genome, therefore it is possible the *flpA* gene lies in the gap region, which would make it 12 of the 18 genomes possessing *flpABD* genes (Table A.6.9). The *flpA* genes encode approximately 384 aa and is homologous to the N-terminal domain of FdhA, whereas the *fdhA* genes encode approximately 712 aa in length. Furthermore, *flpAB* genes are annotated as molybdopterin oxidoreductase in the ZA-10<sup>T</sup>, KM1H5-1P<sup>T</sup> and JMR01 genomes, and since it is not present in the genome of the formate utilising RFM-1<sup>T</sup>, it brings into questions whether the *flpABD* genes are genuinely involved in formate utilisation. The formate transporter gene *fdhC* forms a conserved operon preceding the formate dehydrogenase *fdhAB* genes in all *Methanobrevibacter* genomes analysed, with one or two accessory protein *fdhD* genes present elsewhere in the genome. Directly downstream of *fdhCAB* genes are molybdenum cofactor biosynthesis protein *moaA*, *mobB* genes, followed by the formylmethanofuran dehydrogenase *fwdHFGDBAC* operon in all genomes analysed except RFM-2<sup>T</sup>, in which the *moaA* gene is located elsewhere. To identify whether differences within



the genes contributed to the difference in formate utilisation, *fdhCAB* genes from all *Methanobrevibacter* spp. genomes were aligned (result not shown), and a high degree of gene conservation was observed between the *Methanobrevibacter* genomes. The *fdhA*, *fdhB* and *fdhC* genes have 76%, 72.2% and 67.1% aa identity respectively, but no conservation could be attributed solely to formate utilising genomes. To identify whether differences in regulation may contribute to the difference in formate utilisation, the 5' region of *fdhA* and *fdhC* genes were analysed. The formate utilising genes in *Methanococcus* spp. were upregulated by the absence of H<sub>2</sub> (Wood *et al.* 2003), and perhaps the *fdhCAB* operon in the strains that were found to not utilise formate were down-regulated strongly by the presence of H<sub>2</sub>. The experimental conditions carried out by strains SH<sup>T</sup>, HO<sup>T</sup>, CW<sup>T</sup>, RFM-2<sup>T</sup>, RFM-3<sup>T</sup> and ATM<sup>T</sup> all used H<sub>2</sub>:CO<sub>2</sub> (80:20, v/v, 202 kPa), therefore reducing the partial pressure of H<sub>2</sub> in culture might enable formate utilisation of these strains. The genes involved in proline, tryptophan and coenzyme M biosynthesis are absent from the core genome. All genomes analysed lack the genes required to produce proline, therefore *Methanobrevibacter* spp. may require external proline to survive. Most strains other than PS<sup>T</sup> and KM1H5-1P<sup>T</sup> are reported to require amino acid containing supplements such as yeast extract, casamino acid, or fecal extract to enable good growth. Perhaps PS<sup>T</sup> and KM1H5-1P<sup>T</sup> possess an unknown pathway of proline biosynthesis. JMR01 lacks the genes required to produce tryptophan, and may require external supply to survive, which is conceivably found in the yeast extract and human fecal extract supplemented to JMR01 under culture conditions (Ferrari *et al.* 1994). D5 lacks the *thrB* gene required to produce threonine, and may require external supply of threonine to survive. A similar conundrum to the formate utilisation is observed in the biosynthesis of CoM. M1<sup>T</sup> has long been known to require external CoM for growth (Taylor *et al.* 1974), and the genome sequence showed the lack of *comADE* genes required for CoM biosynthesis (Leahy *et al.* 2010). Two other *Methanobrevibacter* spp. were also shown to require external CoM supply in culture, JH1<sup>T</sup> and SH<sup>T</sup> of the *Mbb. wolinii* clade. However, both genomes are predicted to encode the full set of genes required for CoM biosynthesis. The genes *comABCDE* were aligned between all genomes analysed, the genes were highly conserved, with an average of 67.7%, 53.3%, 66%, 69.2% and 62.4% aa identity respectively. There are minor difference in amino acid residues, but whether this affects the functionality of the enzyme would require experimental evidence. In ComE, the residue 25 is F and residue 33 is I in all genomes but SH<sup>T</sup>, JH1<sup>T</sup> and AbM4 which used V for both residues, residue 89 is P in all genomes except SH<sup>T</sup>, which has a C at residue 89. In ComC, the residue 42 is T in all genomes except SH<sup>T</sup>, which has a S at residue 42. In ComB, the residue 97 is T in all genomes except SH<sup>T</sup>, which

has a S at residue 97, the residue 103 is I in all genomes except SH<sup>T</sup> and M1<sup>T</sup>, which has V and T respectively at residue 103, the ComB is also 10 to eight residues shorter at the C-terminal in M1<sup>T</sup>, SH<sup>T</sup>, JH1<sup>T</sup> and AbM4 genomes. Whether or not these sequence differences could lead to difference in functionality requires further investigation.

There are two gene families identified as being conserved amongst the *Methanobrevibacter* genomes of rumen origin (Table 6.15). The glycosyltransferase GT2 family protein is classified under cell wall/membrane biogenesis in COG categories, it is likely involved in protein glycosylation and may be an adaptation to the rumen environment.

There are 15 gene families identified as conserved amongst the genomes of *Mbb. gottschalkii* clade (Table 6.16). Although most of the gene families lack predicted function, the aforementioned putative formate dehydrogenase gene, *flpA*, and a MFS (Major facilitator superfamily) transporter gene family consisting of two genes are functionally characterised and conserved to all members of *Mbb. gottschalkii* clade. The MFS transporters are known to catalyse the transport of a broad range of solutes (Kaback *et al.* 2001), it may be of interest to identify the substrate of this particular MFS transporter gene family experimentally.

There are 171 gene families uniquely conserved in the genomes of the *Mbb. ruminantium* clade (Figure 6.16). While 125 gene families are poorly characterised, there are 17 gene families involved in coenzyme transport and metabolism (Table 6.14). It was found only members of the *Mbb. ruminantium* clade carry the full complement of genes *bioWFADB* required to produce biotin from pimelate (Bower *et al.* 1996), which suggests only members of *Mbb. ruminantium* clade can carry out *de novo* biosynthesis of biotin. All the genomes analysed are predicted to encode *pycAB* genes encoding a biotin dependent pyruvate carboxylase along with *birA* gene encoding a biotin-acetyl-CoA carboxylase ligase, which suggests *Methanobrevibacter* spp. other than YLM1, M1<sup>T</sup> and KM1H5-1P<sup>T</sup> would require an external supply of biotin to survive. The biotin transporter BioY that could facilitate biotin uptake has been predicted in ten of the genomes analysed, including D5, ZA-10<sup>T</sup>, SM9, YE315 and HO<sup>T</sup> of *Mbb. gottschalkii* clade, JMR01, PS<sup>T</sup>, AbM4, SH<sup>T</sup>, RFM-1<sup>T</sup>, RFM-2<sup>T</sup>, RFM-3<sup>T</sup>, ANOR1 and ATM<sup>T</sup>, which suggests these organisms are capable of biotin uptake.

There are 168 gene families uniquely conserved in the genomes of the *Mbb. wolinii* clade (Figure 6.16), while 120 gene families are poorly characterised, four gene families are classified under cellwall/membrane/envelope biogenesis (Table 6.14). Most notable is a gene encoding aspartate racemase involved in conversion of L-aspartate to D-aspartate (Schleifer

and Kandler 1972), this gene is not predicted outside of the *Mbb. wolinii* clade. Bacteria utilise D-aspartate ligase to add D-aspartate onto peptidoglycan pentapeptide (Bellais *et al.* 2006), no homolog of D-aspartate ligase gene were identified in all of the *Methanobrevibacter* spp. genomes analysed, which makes it uncertain whether the D-aspartate is attached to the pseudomurein. Possibly the enzymes involved differ strongly to the bacterial counterpart that targets peptidoglycan. The possible incorporation of D-aspartate into the pseudomurein contradicts the dogma that pseudomurein lacks D-amino acids and muramic acid (Konig and Kandler 1979). 2% (D/D+L) D-aspartate has been identified in the membrane fractions of *Pyrobaculum islandicum*, *Methanosarcina barkeri* and *Halobacterium salinarium* (Nagata *et al.* 1999), which suggests the cell membrane of members of the *Mbb. wolinii* clade may contain D-aspartate.

There are 444 gene families identified as unique to D5 (Table A.6.04), including six genes encoding glycosyl transferases that may be involved in exopolysaccharide formation, and 75 adhesin-like proteins. The adhesin-like proteins may play a role in adherence to solid substrates, host cells or other microorganisms within the rumen. A set of genes unique to D5 were the endonuclease (D5\_2164) and methylase (D5\_2165) of a type III restriction system, while type I and type II restriction systems are found in other *Methanobrevibacter* genomes. Restriction systems are considered a defence system against phage and plasmids, the host DNA is normally methylated by methylases post-replication, and restriction endonuclease targets non-methylated DNA, such as phage and plasmids (Tock and Dryden 2005). This suggests D5 may have an additional line of defence against foreign genetic elements.

Other noteworthy genes unique to D5 includes a non-ribosomal surfactin synthetase (D5\_0482). As mentioned in Chapter 4, Section 4.3, NRPSs are responsible for biosynthesis of small peptides and antibiotics (Witting and Sussmuth 2011), that are produced by consecutive condensation of substrates (Caboche *et al.* 2008). Surfactin refers to bacterial cyclic lipopeptide consisting of hydrophobic fatty acids and hydrophilic peptide chains (Roongsawang *et al.* 2010). The first surfactin was discovered in *Bacillus subtilis* (Arima *et al.* 1968), and it was found to have antibacterial and antitumor activities (Tsukagoshi *et al.* 1970; Kameda *et al.* 1974). Keeping a gene as large as a NRPS in the genome is energetically expensive, therefore it seems likely that the gene provides certain advantages for D5 survival within the rumen. D5 is not the only member of *Methanobrevibacter* that possess NRPS genes, M1<sup>T</sup> and SM9 have also been predicted to contain NRPSs (Leahy *et al.* 2013; Kelly *et al.* 2016). Additionally, YE315, ZA-10<sup>T</sup>, RFM-1<sup>T</sup> and ANOR1 are also predicted to possess NRPSs.

All *Methanobrevibacter* genomes analysed are predicted to possess a full complement of genes required to carry out hydrogenotrophic methanogenesis from CO<sub>2</sub> and H<sub>2</sub> (Table A.6.9). The C<sub>1</sub> group is transferred from CO<sub>2</sub> to MF by Fwd, which acquires electrons from reduced ferredoxin generated by the Eha complex via a Na<sup>+</sup> potential (de Poorter *et al.* 2003; Thauer *et al.* 2010). The Ehb complex is also involved in the supply of reducing potential, however the reducing potential is supplied to the autotrophic CO<sub>2</sub> assimilation in central carbon metabolism rather than the methanogenesis pathway (Major *et al.* 2010). The C<sub>1</sub> group is then transferred to H<sub>4</sub>MPT by Ftr, and reduced sequentially from formyl to methenyl by Mch, from methenyl to methylene by Mtd or Hmd, from methylene to methyl by Mer (Thauer 1998; Shima *et al.* 2002). There are two enzymes that can reduce methenyl-H<sub>4</sub>MPT to methylene-H<sub>4</sub>MPT, the F<sub>420</sub> dependent Mtd and the F<sub>420</sub> independent Hmd. 13 *Methanobrevibacter* genomes are predicted to encode both enzymes. ANOR1, RFM-1<sup>T</sup>, RFM-2<sup>T</sup>, RFM-3<sup>T</sup> and ATM<sup>T</sup> are predicted to only encode Mtd, however hydrogenotrophic methanogenesis can be carried out by either Mtd or Hmd (Hendrickson and Leigh 2008). While most of the genes encoding hydrogenotrophic methanogenesis exist in single copy, additional copies of the *mtd* gene are predicted in the D5, ZA-10<sup>T</sup> and SM9 genomes, an additional copy of *hmd* gene is predicted in the YE315, ZA-10<sup>T</sup> and SM9 genomes, and an additional copy of *mch* gene is predicted in the JMR01 genome. The multiple copies of these genes suggests their functional importance to the methanogen hosts. The H<sub>2</sub> and F<sub>420</sub>H<sub>2</sub> can be interconverted by the Hmd-Mtd cycle (Hendrickson and Leigh 2008), which suggests ANOR1, RFM-1<sup>T</sup>, RFM-2<sup>T</sup>, RFM-3<sup>T</sup> and ATM<sup>T</sup> would not be able to utilise the Hmd-Mtd cycle to interconvert F<sub>420</sub>H<sub>2</sub> and H<sub>2</sub>. It is also possible that the *hmd* gene may be present in the gap region of ANOR1, RFM-1<sup>T</sup>, RFM-2<sup>T</sup>, RFM-3<sup>T</sup> and ATM<sup>T</sup>, as these genomes are currently in draft form, only completion of the genome can rule out this possibility. Both Mtd and Mer require reducing potential supplied from F<sub>420</sub>H<sub>2</sub> (Thauer 1998; Shima *et al.* 2002). All genomes are predicted to encode the genes required for F<sub>420</sub> biosynthesis and the Frh hydrogenase (Table A.6.4). Alcohol may act as an alternative source of reducing potential for F<sub>420</sub> in the presence of the F<sub>420</sub>-dependent NADP reductase NdpG and NADP-dependent alcohol dehydrogenase (Berk and Thauer 1997), as demonstrated in M1<sup>T</sup> (Leahy *et al.* 2010). While all genomes are predicted to encode the *npdG* gene, the homologous gene to the NADP-dependent alcohol dehydrogenase is only predicted in YE315 and three members of the *Mbb. ruminantium* clade. The conservation of the gene encoding NADP-dependent alcohol dehydrogenase in the *Mbb. ruminantium* clade suggests it could be a conserved characteristic of this group. The conservation of the *ndpG* gene suggests

other *Methanobrevibacter* spp. that lack the NADP-dependent alcohol dehydrogenase may utilise an unknown source of reducing potential other than alcohol.

The methyl group is transferred to CoM by Mtr, which is the main  $\text{Na}^+$  potential generating enzyme complex (Harms *et al.* 1995). The  $\text{Na}^+$  potential is utilised by the  $\text{A}_1\text{A}_0$  ATP to generate ATP, however the  $\text{A}_1\text{A}_0$  ATP synthase can also use a  $\text{H}^+$  gradient at low pH and low  $\text{Na}^+$  condition (McMillan *et al.* 2011). The methyl-CoM is then reduced to  $\text{CH}_4$  by the Mcr/Mrt. complex. Isozymes Mcr and Mrt display different operon structures within a genome, *mcrBDCGA* and *mrtBDGA* respectively (Friedrich 2005). The two isozymes were found to be differentially regulated due to availability of  $\text{CO}_2$  and  $\text{H}_2$ . Mrt is favoured in high  $\text{H}_2$  availability, whereas Mcr is favoured in low  $\text{H}_2$  availability (Bonacker *et al.* 1993). M1<sup>T</sup> and AbM4 were found to only contain the *mcr* genes, and thus may favour a low  $\text{H}_2$  environment (Leahy *et al.* 2010; Leahy *et al.* 2013). In addition, ovine rumen metatranscriptomics has revealed Mcr encoding genes to have a significantly higher expression in high  $\text{CH}_4$  emitting animals, while no significant difference was found for the expression of Mrt encoding genes between high and low  $\text{CH}_4$  emitting animals (Shi *et al.* 2014). In this study, YE315, ANOR1, SH<sup>T</sup> and JH1<sup>T</sup> are predicted to only carry the *mcr* genes, similar to M1<sup>T</sup> and AbM4, which suggests these *Methanobrevibacter* spp. may be similar to M1<sup>T</sup> and AbM4 and favour a low  $\text{H}_2$  environment. All three members of the *Mbb. wolinii* clade were found to only possess *mcr*, which suggests it could be a conserved feature of the *Mbb. wolinii* clade. Conversely, other *Methanobrevibacter* spp. genomes analysed encodes both *mcr* and *mrt*, which suggests they may be more versatile and capable of growing in a wider range of  $\text{H}_2$  partial pressures.

*Methanobrevibacter* spp. are the predominant members of rumen methanogens, but it is not the only genus of the order Methanobacteriales in the rumen (Janssen and Kirs 2008; Seedorf *et al.* 2014; Henderson *et al.* 2015; Seedorf *et al.* 2015). There are several difference between *Methanobrevibacter* spp. and other methanogens that may have contributed to its predominance. Both *Methanobrevibacter* spp. and *Methanosphaera* spp. belong to the Methanobacteriales order, but they differ morphologically and metabolically (Fricke *et al.* 2006). *Methanobrevibacter* spp. are short rods while *Methanosphaera* spp. are spherical shaped, *Methanobrevibacter* spp. rely on hydrogenotrophic methanogenesis to live, while *Methanosphaera* rely on methylotrophic methanogenesis to survive (Fricke *et al.* 2006). Another order of methanogens that possesses the hydrogenotrophic methanogenesis pathway are the Methanosarcinales, but because Methanosarcinales contain cytochromes (Heiden *et al.* 1994; Deppenmeier 2004; Maeder *et al.* 2006), the hydrogenotrophic methanogenesis would

require much higher concentration of H<sub>2</sub> in the environment for the reaction to be thermodynamically favorable (Thauer *et al.* 2008). Therefore, Methanosarcinales primarily rely on the disproportionating methylotrophic methanogenesis to survive in the rumen (Hippe *et al.* 1979; Maeder *et al.* 2006; Liu and Whitman 2008). The conservation of hydrogenotrophic methanogenesis genes across all *Methanobrevibacter* genomes suggests *Methanobrevibacter* spp. occupy an important niche in the rumen. Hydrogenotrophic methanogenesis requires four moles of H<sub>2</sub> to produce one mole of CH<sub>4</sub>, and it is able to effectively deplete H<sub>2</sub> and prevent its accumulation in the rumen, which is beneficial to the fermentative microbes that form a syntrophic relationship with the *Methanobrevibacter* spp.

## 6.4 Conclusion

Members of the *Methanobrevibacter* genus are the predominant rumen methanogen, and the *Mbb. gottschalkii* clade is the most dominant within the *Methanobrevibacter* (Janssen and Kirs 2008; Henderson *et al.* 2015). *Methanobrevibacter* sp. D5 is the fifth member of the *Mbb. gottschalkii* clade to be genome sequenced, alongside the genomes of *Mbb. millerae* ZA-10<sup>T</sup>, *Mbb. millerae* SM9, *Methanobrevibacter* sp. YE315 and *Mbb. gottschalkii* HO<sup>T</sup>. The genes present in the D5 genome predicts it utilises CO<sub>2</sub>, H<sub>2</sub> and formate for hydrogenotrophic methanogenesis, acetate as a carbon source, ammonium as a nitrogen source, and potentially requires an external supply of threonine and proline to survive. The utilisation of formate would require experimental evidence to prove, as all 18 *Methanobrevibacter* genomes analysed possess the full set of genes required to utilise formate, yet only eight type strains have been observed to be capable of doing so.

This study revealed the core genome conserved between 18 *Methanobrevibacter* spp and also clade-specific features. The core genome includes the full complement of genes required for methanogenesis, with only the Mcr, but not Mrt, complex conserved across *Methanobrevibacter* spp. This study also revealed clade specific genes, and members of the *Mbb. ruminantium* clade may be specialised in biotin biosynthesis and the use of alcohol as a source of reducing potential. Members of the *Mbb. wolinii* clade may have a specialised cell envelope that employs D-aspartate, as well as a lack of an Mrt complex that might make them less versatile than other *Methanobrevibacter* spp. Members of the *Mbb. gottschalkii* clade carries the lowest amount of clade specific genes, mostly with unknown function. The D5 genome was found to possess unique genes including large number of adhesins genes and a

large NRPS gene predicted to produce surfactin, which may be beneficial to the survival of D5 in the rumen.

The *Mbb. gottschalkii* clade and *Mbb. ruminantium* clade are the two most dominant clades of rumen methanogens, this study finds the methanogenesis pathway of all *Methanobrevibacter* spp. genomes analysed to be highly conserved with only minor differences, which is unlikely to explain the co-existence of *Methanobrevibacter* spp. within the rumen. The most probable explanation to the co-existence of *Methanobrevibacter* spp. in the rumen is the large number of genes involved in adhesion. By binding to different microorganisms, digesta or host surface, the *Methanobrevibacter* spp. may be able to establish their own specialised niche despite their similarity in energy metabolism.

## Chapter 7

### Summary, conclusion and future direction

#### 7.1 Rationale

New Zealand has a large number of ruminant livestock relative to its human population, and the CH<sub>4</sub> produced by these animals via enteric fermentation accounts for 79.9% of New Zealand's total CH<sub>4</sub> emissions (Ministry for the environment [MfE], 2015). In order to formulate strategies for enteric CH<sub>4</sub> mitigation, it is necessary to understand the source of CH<sub>4</sub> and its formation in the rumen environment.

Rumen fermentation produces H<sub>2</sub> as a by-product, which is mainly utilised by rumen methanogens (Hungate 1966). Because most rumen methanogens use H<sub>2</sub>, methanogen activity and growth are closely linked to the activity of H<sub>2</sub>-producing microbes, such as rumen protozoa, fungi and bacteria (Janssen 2010). Understanding the characteristics of rumen methanogens and their complex relationships with other rumen microbes, is therefore essential to formulate strategies to mitigate enteric CH<sub>4</sub> emissions. Mitigation strategies such as ruminant vaccination and chemical inhibitors of methanogens requires selective targets, which ideally are functions that are conserved across all rumen methanogens and absent from the animal host and other members of the rumen microbiome (Attwood *et al.* 2008). The genome sequences of rumen methanogens can effectively identify the genes encoding these conserved functions and also improves our understanding of rumen methanogen biology. Before this PhD project began, very little was known about the genomics of rumen methanogens and only one rumen methanogen - *Mbb. ruminantium* M1<sup>T</sup>, had been genome sequenced (Leahy *et al* 2010). The order Methanomassiliicoccales had only recently been defined, and although there were enrichment cultures available for Rumen Cluster C organisms, there were no genome sequences available for rumen representatives of this order. The aim of the thesis was to i) to determine the genome composition of the methanogenic archaeon RCC ISO4-H5, and how this differs from other sequenced methanogens, and ii) determine the metabolic scheme of RCC ISO4-H5, how it grows in the rumen, and use these features to attempt isolation of a pure culture of this organism, iii) sequence the genome of an additional representative of the genus *Methanobrevibacter* (*Methanobrevibacter* sp. D5) to look at genome composition variation within this genus and to determine how these differences allow co-existence of multiple *Methanobrevibacter* species in the rumen.



## 7.2 Advancements since the start of this thesis.

The improvement of high throughput sequencing technologies and the corresponding decrease in the cost of genome sequencing, has led to a dramatic increase in the amount of microbial genome sequence data. Since my PhD project began, 10 additional Methanomassiliicoccales genomes have become available, including “*Candidatus* Methanomassiliicoccus intestinalis Mx1-Issoire”, “*Candidatus* Methanomethylophilus alvus Mx1201” and *Methanomassiliicoccus luminyensis* B10 sequenced from human sources (Borrel *et al.* 2012; Dridi *et al.* 2012; Borrel *et al.* 2013), “*Candidatus* Methanoplasma termitum MpT1” sequenced from the termite gut (Lang *et al.* 2015), Thermoplasmatales archaeon BRNA1 (NCBI Reference Sequence NC\_020892.1), “*Candidatus* Methanomethylophilus” sp. 1R26, Methanomassiliicoccales archaeon RumEn M1 and RumEn M2 sequenced from the bovine rumen (Noel *et al.* 2016; Sollinger *et al.* 2016), and Methanogenic archaeon ISO4-G1 and ISO4-G11 sequenced from the ovine rumen (Kelly *et al.* 2016; Leahy *et al.* Unpublished).

Similar genome sequencing advancements amongst *Methanobrevibacter* spp. have been made as part of the Hungate1000 project (Creevey *et al.* 2014), including *Methanobrevibacter* sp. YE315, *Mbb. millerae* ZA-10<sup>T</sup> from the bovine rumen, *Mbb. gottschalkii* HO<sup>T</sup> from horse faeces, *Mbb. wolinii* SH<sup>T</sup> from sheep faeces, and *Mbb. olleyae* KM1H5-1P<sup>T</sup> from the ovine rumen. Additionally, genomes of *Methanobrevibacter* sp. AbM4 (ovine rumen; Leahy *et al.* 2013) and *Mbb. millerae* SM9 (ovine rumen; Kelly *et al.* 2016), *Mbb. olleyae* YLM1 (ovine rumen; Kelly *et al.* 2016), *Mbb. boviskoreani* JH1<sup>T</sup> (bovine rumen; Lee *et al.* 2013), *Mbb. oralis* JMR01 (human faeces; Khelaifia *et al.* 2014) and *Mbb. arboriphilus* ANOR1 (human faeces; Khelaifia *et al.* 2014; Khelaifia *et al.* 2014), *Mbb. cuticularis* RFM-1<sup>T</sup> (termite gut; NCBI whole genome shotgun sequencing project LWMW000000000), *Mbb. curvatus* RFM-2<sup>T</sup> (termite gut; NCBI whole genome shotgun sequencing project LWMV000000000), *Mbb. filiformis* RFM-3<sup>T</sup> (termite gut; NCBI whole genome shotgun sequencing project LWMT000000000) and *Mbb. acididurans* ATM<sup>T</sup> (anaerobic fermenter, Rosewarne *et al.*, unpublished) are available. These new sequences have enhanced the comparative analyses conducted in this thesis, and have allowed conserved traits and important metabolic characteristics of the genus *Methanobrevibacter* and the order Methanomassiliicoccales to be identified with greater reliability.

### 7.3 Summary of thesis results

A large part of this thesis has been the determination of the genome composition of methanogenic archaeon ISO4-H5 and its comparison to available genomes of members of the order Methanomassiliicoccales. The ISO4-H5 genome sequence revealed it as a hydrogen-dependent methylotrophic methanogen. Its predicted substrates for methanogenesis include mono-, di-, and tri-methylamine, methanol and the methylated thiol compound, methyl-3-methylthiopropionate (M3MSP). Experimental verification of these substrates confirmed the utilisation of these substrates. The comparative genomic analyses of members of the Methanomassiliicoccales revealed the absence of the hydrogenotrophic methanogenesis pathway prior to the methyl CoM reduction step catalysed by the Mcr/Mrt enzyme. This reaffirms the inability of Methanomassiliicoccales to disproportionate the methyl-substrates to CO<sub>2</sub> and their reliance on methyl-compound reduction only. All the genomes analysed are predicted to contain the genes required to allow utilisation of methylamine as a substrate for methylotrophic methanogenesis. All of the Methanomassiliicoccales genomes contain the genes which are predicted to encode autotrophic CO<sub>2</sub> fixation through AMP recycling, catalysed by a RubisCo operon. The alternative coding of the amber stop codon to encode for pyrrolysine incorporation is a conserved feature found in all the Methanomassiliicoccales genomes analysed. However, the Mx1, B10 and RumEn M1 organisms possess large numbers of genes with genuine amber stop codons, and appear to regulate the incorporation of pyrrolysine only where the amber codons are associated with methylamine methyltransferase genes. The remaining Methanomassiliicoccales organisms appear to have fewer amber codons in their genomes, but appear not to regulate the incorporation of pyrrolysine and are predicted to incorporate pyrrolysine at a higher frequency. The capability of CoM biosynthesis within the Methanomassiliicoccales appeared to be environment-dependent, as all three members of Methanomassiliicoccales from human sources are capable of CoM biosynthesis, whereas none of the rumen- or termite-derived strains possess genes required for CoM biosynthesis.

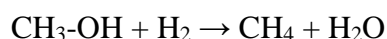
This study successfully isolated ISO4-H5 from its original enrichment culture, and is the first pure culture of a ruminal member of the order Methanomassiliicoccales. ISO4-H5 was identified as consuming ethanol, methanol and nicotinamide via NMR analysis of spent culture supernatants. Transcriptomic analyses were also conducted and identified specific set of ISO4-H5 genes that were differentially regulated in response to high and low H<sub>2</sub> environments. In the low H<sub>2</sub> conditions, several genes involved in methanogenesis and supply of reducing

potential were up-regulated, including *mrtA*, *mvhAGD* and *hdrABC* genes. All genomes analysed possess a small hypothetical protein (AR505\_1068, and corresponding homologues) adjacent to the *mrtA* gene, which in ISO4-H5 was regulated in a manner similar to *mrtA*, which suggests it could be associated with the Mrt complex, or the process of methyl CoM reduction. This study also revealed that the genes encoding methylamine and M3MSP utilisation were up-regulated by the presence of methylamine.

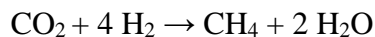
The analyses of *Methanobrevibacter* genomes revealed that the D5 genome is the second largest amongst the 18 available *Methanobrevibacter* spp. genomes analysed. D5 is predicted to utilise CO<sub>2</sub>, H<sub>2</sub> and formate for hydrogenotrophic methanogenesis. The comparison between *Methanobrevibacter* spp. genomes revealed their genomes are highly conserved, with a “core” genome of 764 gene families, including all the genes involved in hydrogenotrophic methanogenesis and general housekeeping functions. A large number of genes encoding adhesin-like proteins and a NRPS were unique to D5. Therefore, it is hypothesised that differences in microbial interactions and the type of microbial partner may be an important clue to explain the coexistence of many different *Methanobrevibacter* spp. in the rumen. Furthermore, a number of *Methanobrevibacter* clade-specific genetic traits were identified in this study. Only members of the *Mbb. ruminantium* clade encode biotin biosynthesis, which suggests other *Methanobrevibacter* spp. acquire exogenous biotin via cross-feeding. Only members of the *Mbb. wolinii* clade have an aspartate racemase gene to produce D-aspartate, which suggests the cell envelope may contain D-aspartate. *Methanobrevibacter* strains M1<sup>T</sup>, AbM4, YE315, SH<sup>T</sup>, JH1<sup>T</sup>, RFM-1<sup>T</sup>, RFM-2<sup>T</sup>, RFM-3<sup>T</sup>, ATM<sup>T</sup> and ANOR1 are predicted to only encode Mcr, and not Mrt complex, which may indicate a difference in regard to how they respond to different levels of H<sub>2</sub> and their effective H<sub>2</sub> thresholds.

#### **7.4 Relevance of thesis findings to methane formation in the rumen**

The genome sequences generated and analysed in this thesis provided insights into the lifestyle of two very different rumen methanogens, a member of the Methanomassiliicoccales, methanogenic archaeon ISO4-H5 and a member of the *Methanobrevibacter* genus, *Methanobrevibacter* sp. D5. These methanogens have very different modes of methanogenesis which impacts on how they live and survive in the rumen. ISO4-H5 depends strictly on methylotrophic methanogenesis and is unable to disproportionate methyl-substrates. Using methanol as an example of a methyl-substrate, the theoretical standard free energy change ( $\Delta G^{\circ}$ ) from the reaction:



is -112.5 kJ/mole (Thauer *et al.* 1977). On the other hand, D5 relies on hydrogenotrophic methanogenesis according to the following reaction:



The  $\Delta G^{\circ'}$  from hydrogenotrophic methanogenesis is -131 kJ/mole (Thauer *et al.* 2008). Based on these stoichiometries, D5 and other hydrogenotrophic methanogens, require four moles of  $\text{H}_2$  to produce one mole of  $\text{CH}_4$ , whereas ISO4-H5-like Methanomassiliicoccales only require one mole of  $\text{H}_2$  to produce one mole of  $\text{CH}_4$ . This suggests that the Methanomassiliicoccales have a lower threshold of dissolved  $\text{H}_2$  than the hydrogenotrophic methanogens. For methanogens to grow, the difference in free energy generated from methanogenesis needs to exceed the phosphorylation potential of  $\text{ADP} + \text{P}_i \rightarrow \text{ATP}$  (Thauer *et al.* 2008). The free energy differential for methanogenesis also depends on the dissolved  $\text{H}_2$  concentration (Janssen 2010). The  $\text{H}_2$  in the rumen is produced by microbial fermentation of feed, it is dynamic and influenced by multiple factors that are interconnected, such as feed type, feed passage rate, ruminal pH, and the relative ratio of acetate and propionate production (Janssen 2010). Efficient removal of  $\text{H}_2$  increases the rate of fermentation by eliminating the negative feedback inhibition imposed by  $\text{H}_2$  (McAllister and Newbold 2008; Janssen 2010). The free energy change ( $\Delta G'$ ) for ISO4-H5 and D5 under varying  $\text{H}_2$  concentrations can be calculated using the following equations:

$$\text{D5: } \Delta G' = \Delta G^{\circ'} + 2.3RT \log [\text{CH}_4]/[\text{H}_2]^4 \times [\text{CO}_2]$$

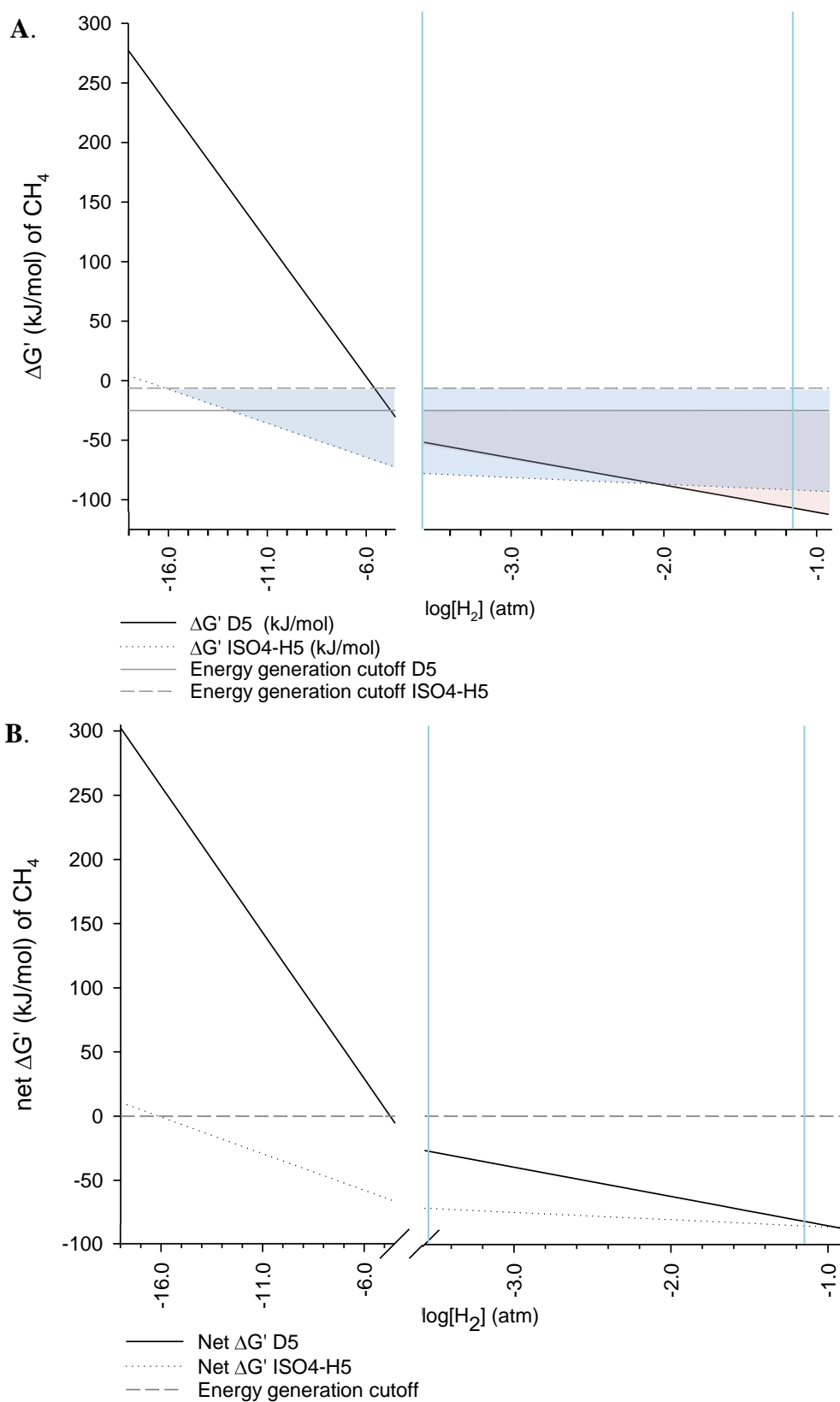
$$\text{ISO4-H5: } \Delta G' = \Delta G^{\circ'} + 2.3RT \log [\text{CH}_4]/[\text{H}_2] \times [\text{CH}_3\text{OH}]$$

Where R is the gas constant  $8.314 \text{ JK}^{-1} \text{ mol}^{-1}$ , T is absolute temperature  $298.15^\circ \text{K}$ .

Under standard conditions, the methylotrophic methanogenesis in ISO4-H5 only needs one mole of  $\text{H}_2$  to produce one mole of  $\text{CH}_4$ , and pumps one mole of ions every two moles of  $\text{CH}_4$  due to electron bifurcation. The hydrogenotrophic methanogenesis in D5 needs four moles of  $\text{H}_2$  to produce one mole of  $\text{CH}_4$ , and pumps two moles of ions protons for every mole of  $\text{CH}_4$  produced (Thauer *et al.* 2008). Four moles of ions protons are required to drive one mole of ATP production. Assuming the energy required to make one mole of ATP is 50 kJ, ISO4-H5 diverts 6.25 kJ of free energy change to ATP synthesis for every mole of  $\text{CH}_4$  produced, while D5 diverts 25 kJ of free energy change to ATP synthesis for every mole of  $\text{CH}_4$  produced. If

the free energies are plotted against the log [H<sub>2</sub>] (Figure 7.1), it shows that the concentration of dissolved H<sub>2</sub> typically present in the rumen is sufficient to support ATP generation from methanogenesis in both ISO4-H5 and D5. The higher the dissolved [H<sub>2</sub>] becomes in the rumen, the more energetically favourable (more negative  $\Delta G'$ ) the methanogenesis reaction becomes (Janssen 2010). This diagram also shows that compared to ISO4-H5, D5 has a narrower range of H<sub>2</sub> concentrations in which it can generate energy. (Figure 7.1). Dissolved H<sub>2</sub> can achieve a theoretical maximum of 737  $\mu\text{M}$ , but inside the rumen H<sub>2</sub> is usually between 50  $\mu\text{M}$  (0.067843 atm) and 0.1  $\mu\text{M}$  (0.000268 atm) or lower (Janssen 2010). This suggests that under rumen conditions both ISO4-H5 and D5 are constantly exposed to sufficient concentrations of dissolved H<sub>2</sub> to permit methanogenesis. While D5 is predicted to generate more free energy change per mole of CH<sub>4</sub> than ISO4-H5 at [H<sub>2</sub>] > approximately log[H<sub>2</sub>] of -2, ( $\approx 0.01$  atm, Figure 7.1A), when the energy requirements of ATP biosynthesis is considered, ISO4-H5 always has a greater net free energy change than D5 under rumen conditions (Figure 7.1B). However, *Methanobrevibacter* spp. are the predominant rumen methanogens (Janssen and Kirs 2008; Henderson *et al.* 2015; Seedorf *et al.* 2015) and thus must be able to grow well under the conditions that prevail in the rumen. The fact that members of the Methanomassiliicoccales are not the dominate rumen methanogens also suggests that their growth is limited by factors other than dissolved H<sub>2</sub> and is most likely governed by the availability of methyl substrates.

Despite the prediction that H<sub>2</sub> concentration is unlikely to limit Methanomassiliicoccales growth, transcriptomic analysis revealed up-regulation of ISO4-H5 *hdrABC* and *mvhAGD* genes in the low H<sub>2</sub> environment, presumably as a response to scavenge more H<sub>2</sub>. This suggests that the decrease of  $\Delta G'$  for ATP generation by low [H<sub>2</sub>] is significant to the ISO4-H5 cells, and rather than produce less ATP, ISO4-H5 appears to adapt to the low [H<sub>2</sub>] by up-regulating genes to scavenge H<sub>2</sub>. This shows that ISO4-H5 maximises the gain of energy from the rumen environment under changing conditions and likely reflects the competitive nature of the rumen.



**Figure 7.1 Theoretical  $\Delta G'$  of methanogenesis in relation to ATP synthesis in the rumen.**  $\Delta G^{\circ'}$  of  $\text{CO}_2 + 4 \text{H}_2 \rightarrow \text{CH}_4 + 2 \text{H}_2\text{O}$  is -131 kJ/mole (Thauer *et al.* 2008),  $\Delta G^{\circ'}$  of  $\text{CH}_3\text{-OH} + \text{H}_2 \rightarrow \text{CH}_4 + \text{H}_2\text{O}$  is -112.5 kJ/mole (Thauer *et al.* 1977) under standard condition. Assuming the total gas pressure in the rumen is  $1.01325 \times 10^5$  pa, comprising of  $\text{CO}_2$ : $\text{CH}_4$  of 65.35%:26.76%,  $[\text{CO}_2]$  would be 0.6535 atm,  $[\text{CH}_4]$  would be 0.2676 atm (McArthur

and Miltimore 1961),  $[H_2]$  observed in the rumen is from 0.1  $\mu M$  (0.000268atm,  $\log [H_2] = -3.57$ ) to 50  $\mu M$  (0.067843 atm,  $\log [H_2] = -1.17$ ) as indicated between cyan lines (Janssen 2010).  $[CH_3OH]$  is 8.74E-4 mol/L in the rumen fluid of cows fed 40% hay 60% grain (Vantcheva *et al.* 1970).  $\Delta G'$  of  $ADP + P_i \rightarrow ATP$  is considered to be -50 kJ/mole, D5 generates approximately 0.5 mole of ATP per mole of  $CH_4$ , and ISO4-H5 generates 0.125 mole of ATP per mole of  $CH_4$ . **A.** Free energy change of ISO4-H5 and D5, blue and pink shaded area represents the  $\Delta G'$  region in which ATP synthesis from  $ADP + P_i$  is possible for ISO4-H5 and D5 respectively. **B.** Net free energy change of ISO4-H5 and D5 after ATP biosynthesis is considered.

While ISO4-H5 and D5 showed clear differences in their predicted  $H_2$  thresholds and their different ruminal niches, the difference between methanogens of the same order or genus is less apparent. Most of the differences between Methanomassiliicoccales appear to be associated with genes encoding cell envelope surface features rather than central metabolism. ISO4-H5 is predicted to encode a range of adhesin-like proteins that may mediate binding to a variety of surfaces. It also encodes a cluster of genes involved in exopolysaccharide biosynthesis that is not found in other members of the Methanomassiliicoccales, which may also be involved in cell surface modifications. Similarly, D5 is predicted to encode 50 adhesin-like proteins that are not found in other *Methanobrevibacter* spp. The interaction and binding of methanogens to other microorganisms in the rumen or to particular tissues in the ruminant digestive tract, could be important in the establishment of unique methanogen niches. D5 and a few other *Methanobrevibacter* spp. also encode NRPSs that may produce antimicrobial peptides; if so, these antimicrobial peptides may play a part in competition between similar methanogens, and thus influence the success of the producing species.

## 7.5 Future Perspectives

The genome sequence of members of the order Methanomassiliicoccales and the genus *Methanobrevibacter* have offered insights into their lifestyles in the rumen. However, the information inferred from genomes alone is not the full picture and many factors can influence the behaviour of genes, such as transcriptional and translational regulation, post-translational modification, therefore the genomic information serve as a starting point to investigate genes or metabolic traits of interest. Exactly how these genes bring about the metabolic pathway *in vivo*, may only be verified by experimental investigation. For example, the genomes of *Methanobrevibacter* spp. revealed confounding results inconsistent to published literature in the utilisation of formate and requirement of CoM. All genomes analysed possess the necessary genes to utilise formate, yet the  $SH^T$ ,  $HO^T$ ,  $RFM-2^T$ ,  $RFM-3^T$  and  $ATM^T$  are observed to not utilise formate (Leadbetter and Breznak 1996; Leadbetter *et al.* 1998; Miller and Lin 2002;

Savant *et al.* 2002). The M1<sup>T</sup> genome is the only genome to lack the full complement of genes required to produce CoM and requires external CoM to grow (Taylor *et al.* 1974), however SH<sup>T</sup> also requires external CoM for growth and the growth of JH1<sup>T</sup> is stimulated by CoM supplementation (Miller and Lin 2002; Lee *et al.* 2013). These results highlights the importance of culture based experimental validation, it is clear that findings from bioinformatics approaches require biological validation to be meaningful.

Methanogen growth requirements should be conducted in pure cultures, without the uncertainties posed by other microorganisms within the enrichment culture. To date, ISO4-H5 and ISO4-G1 are the only members of the rumen Methanomassiliicoccales for which pure cultures have been obtained. This suggests there should be a renewed attempt to isolate pure cultures of new rumen methanogens, as well as enrichment of previously unculturable members of the Methanomassiliicoccales. Furthermore, many available genome sequences remain in draft status, and completing the genomes by gap closure would reduce the uncertainty posed by a draft genome. Furthermore, the available genome sequences only represent a small proportion of the rumen Methanomassiliicoccales population (Seedorf *et al.* 2014), and in the future, a high resolution metagenome sequencing of the rumen microbiome may enable genome assembly of previously uncultivated rumen methanogens.

While the theoretical calculations of  $\Delta G'$  is useful for defining the thermodynamic boundaries of methanogen growth in the rumen, they are limited to the assumptions applied, and cannot fully portray the actual exchanges that are likely to occur between rumen methanogens and their co-resident microbes. For example, it does not account for factors such as high local H<sub>2</sub> concentrations resulting from methanogens adherence to H<sub>2</sub> producers or the possibility of interspecies electron transfer. Within the range of dissolved H<sub>2</sub> observed in the rumen, thermodynamics favours methane formation from methyl compounds and ISO4-H5 is predicted to be better able to carry out ATP generation than D5 (Figure 7.1). However, *Methanobrevibacter* spp., mainly the *Mbb. gottschalkii* clade, are the predominant methanogens and not members of the order Methanomassiliicoccales. The ruminal level of methyl substrates, such as methanol and methylamines, may limit the proliferation of Methanomassiliicoccales in the rumen. If this is true, feed supplementation with methyl-substrates may elevate the rumen Methanomassiliicoccales population to become the predominant rumen methanogen. Methanol arises in the rumen from cleavage of methylated compounds such as pectin (Siragusa *et al.* 1988), The *in vitro* supplementation of low esterified citrus pectin reduced the methane production by 19%, while high esterified citrus pectin



increased the methane production by 7% (Geerkens *et al.* 2013). Methylamines originate from breakdown of methylated nitrogen compounds such as betaine and choline (Neill *et al.* 1978). Radiolabeled betaine supplementation *in vivo* gave rise to trimethylamine, CH<sub>4</sub> and CO<sub>2</sub> (Mitchell *et al.* 1979), but the methanogen populations were not assessed in these studies. The Tibetan yak was found to have 80.9% of its rumen methanogen population as members of the Methanomassiliicoccales (Huang *et al.* 2012), and while this was mainly attributed from the high tannin level in the alpine *Kobresia* pasture, the level of methyl-substrate from this forage has not been determined and may also contribute to the high Methanomassiliicoccales population. The logical next step to assess whether additional methyl-substrates stimulate Methanomassiliicoccales would be an *in vivo* trial involving prolonged feeding of methylamine compounds followed by assessment of methanogen populations. Assuming all other growth requirements are fulfilled, the hypothesis would be that the rumen Methanomassiliicoccales should proliferate to higher levels than the *Methanobrevibacter* spp.

The *in vitro* growth of ISO4-H5, ISO4-G1 and ISO4-G11 are all very slow, even with excess methanol, and they all produce comparatively low amounts of CH<sub>4</sub> compared to the *Methanobrevibacter* spp. Although this study showed ISO4-H5 utilised external ethanol and nicotinamide, it is obvious that there are additional nutritional requirements of rumen Methanomassiliicoccales that remain to be identified, therefore a defined medium which allows the growth requirements of Methanomassiliicoccales to be determined, would be useful in this regard.

A strong association between *Succinivibrionaceae* and *Methanomassiliicoccaceae* was found in a global census of the rumen microbiome (Henderson *et al.* 2015). This association may be explained, at least partially, by the release of methanol by *Succinivibrio* spp. from the degradation of pectin (Dehority 1969), but this study has demonstrated that methanol is not the limiting condition *in vitro*. Therefore, the association of *Succinivibrionaceae* with *Methanomassiliicoccaceae* deserves further investigation, to better understand the nature of this association and define the metabolites and/or interactions that underlie this relationship. Transcriptomic analyses of *Succinivibrio dextrinosolvens* H5 in pure and co-culture studies with Methanomassiliicoccales would be useful in this regard.

Despite the large differences between members of Methanomassiliicoccales and *Methanobrevibacter* spp, the central Mcr/Mrt and CoM-dependent methyl group reduction remains strongly conserved, as is found in all known methanogens to date. The rumen members

of Methanomassiliicoccales depend on exogenous supply of CoM to survive as they are incapable of *de novo* CoM biosynthesis, therefore structural analogues of CoM would still be inhibitory to these methanogens. Previously, the CoM analogue, 2-bromoethanesulfonate (BES), did not show lasting inhibition of methanogens *in vivo* (Immig *et al.* 1996). It is possible that BES caused the inhibition of CoM auxotrophic methanogens only and led to the drop in methane production, but methanogens capable of *de novo* CoM biosynthesis were less susceptible and filled the vacant niche after a short time. It is possible that the biosynthetic pathway of CoM may be a better target for inhibition, especially the ComDE that are found in the Methanomassiliicoccales Mx1, Mx1201 and B10. If the *de novo* biosynthesis of CoM is inhibited, the Methanomassiliicoccales would initially remain unaffected, but as inhibition persisted and exogenous CoM is diluted by flow feed and water through the rumen, methanogens may eventually be inhibited. Persistent inhibition of CoM biosynthesis, if applied together with CoM analogue such as BES, may be synergistic and a more effective inhibition strategy.

While the transcriptome of ISO4-H5 illustrated the adaptation of ISO4-H5 to a low H<sub>2</sub> environment by up-regulation of *mrtA*, *hdr* and *mvh* genes, it also raised more interesting questions. For example, why are the other subunits of the Mrt complex not up-regulated in the same manner (Poulsen *et al.* 2013) and what is the function of the small hypothetical protein encoded adjacent to the *mrtA* gene? Previously the *mcr/mrt* operons from the rumen of sheep have been classified into three distinct groups (Shi *et al.* 2014). The *mrt* operons of ISO4-H5 and members of order Methanomassiliicoccales are distinct to the *mcr* and *mrt* of the Methanobacteriales. These observations suggest that the molecular mechanism of the Mrt complex in the order Methanomassiliicoccales behaves differently and warrants further investigation. The transcriptomic analyses of ISO4-H5 in the co-culture studies also hinted a possibility of direct interaction with *Ruminococcus flavefaciens* FD1. The interspecies interactions and cross-feeding between methanogens and H<sub>2</sub> producers is an important process that strongly influences methane formation and deserves further investigation.

## Reference

- Abken, H. J., M. Tietze, *et al.* (1998). "Isolation and characterization of methanophenazine and function of phenazines in membrane-bound electron transport of *Methanosarcina mazei* Go1." J Bacteriol 180(8): 2027-2032.
- Abu-Qarn, M., S. Yurist-Doutsch, *et al.* (2007). "*Haloferax volcanii* AglB and AglD are involved in *N*-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer." J Mol Biol 374(5): 1224-1236.
- Albers, S. V. and B. H. Meyer (2011). "The archaeal cell envelope." Nat Rev Microbiol 9(6): 414-426.
- Alikhan, N. F., N. K. Petty, *et al.* (2011). "BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons." BMC Genomics 12: 402.
- Alkalaeva, E., B. Eliseev, *et al.* (2009). "Translation termination in pyrrolysine-utilizing archaea." FEBS Lett 583(21): 3455-3460.
- Allison, C. and G. T. Macfarlane (1989). "Influence of pH, nutrient availability, and growth rate on amine production by *Bacteroides fragilis* and *Clostridium perfringens*." Appl Environ Microbiol 55(11): 2894-2898.
- Allocati, N., M. Masulli, *et al.* (2015). "Die for the community: an overview of programmed cell death in bacteria." Cell Death Dis 6: e1609.
- Altermann, E. (2012). "Tracing lifestyle adaptation in prokaryotic genomes." Front Microbiol 3: 48.
- Altermann, E. and T. R. Klaenhammer (2003). "GAMOLA: a new local solution for sequence annotation and analyzing draft and finished prokaryotic genomes." OMICS 7(2): 161-169.
- Altschul, S. F., W. Gish, *et al.* (1990). "Basic Local Alignment Search Tool." J Mol Biol 215(3): 403-410.
- Ameri, A., D. K. Machiah, *et al.* (2007). "A nonstop mutation in the factor (F)X gene of a severely haemorrhagic patient with complete absence of coagulation FX." Thromb Haemost 98(6): 1165-1169.

Anders, S. and W. Huber (2010). "Differential expression analysis for sequence count data." *Genome Biol* 11(10): R106.

Andreesen, J. R. and L. G. Ljungdahl (1974). "Nicotinamide adenine dinucleotide phosphate-dependent formate dehydrogenase from *Clostridium thermoaceticum*: purification and properties." *J Bacteriol* 120(1): 6-14.

Anderson, M. J., K. E. Ellingsen, *et al.* (2006). "Multivariate dispersion as a measure of beta diversity." *Ecol Lett* 9(6): 683-693.

Andrews, S. (2010). "FastQC: A quality control tool for high throughput sequence data." Reference Source.

Arbing, M. A., S. Chan, *et al.* (2012). "Structure of the surface layer of the methanogenic archaean *Methanosarcina acetivorans*." *Proc Natl Acad Sci U S A* 109(29): 11812-11817.

Arima, K., A. Kakinuma, *et al.* (1968). "Surfactin, a crystalline peptidelipid surfactant produced by *Bacillus subtilis*: isolation, characterization and its inhibition of fibrin clot formation." *Biochem Biophys Res Commun* 31(3): 488-494.

Arnesen, L. P. S., A. Fagerlund, *et al.* (2008). "From soil to gut: *Bacillus cereus* and its food poisoning toxins." *FEMS Microbiol Rev* 32(4): 579-606.

Attwood, G. T., W. J. Kelly, *et al.* (2008). "Analysis of the *Methanobrevibacter ruminantium* draft genome: Understanding methanogen biology to inhibit their action in the rumen." *Aust J Exp Agric* 48(1-2): 83-88.

Attwood, G. T., W. J. Kelly, *et al.* (2008). "Application of rumen microbial genome information to livestock systems in the postgenomic era." *Aust J Exp Agric* 48(7): 695-700.

Aung, H. L., D. Dey, *et al.* (2015). "A high-throughput screening assay for identification of inhibitors of the A<sub>1</sub>A<sub>0</sub>-ATP synthase of the rumen methanogen *Methanobrevibacter ruminantium* M1." *J Microbiol Meth* 110: 15-17.

Avissar, Y. J. and S. I. Beale (1989). "Identification of the enzymatic basis for  $\Delta$ -aminolevulinic acid auxotrophy in a *hemA* mutant of *Escherichia coli*." *J Bacteriol* 171(6): 2919-2924.

- Azad, R. K. and J. G. Lawrence (2007). "Detecting laterally transferred genes: use of entropic clustering methods and genome position." *Nucleic Acids Res* 35(14): 4629-4639.
- Bachmann, B. O. and J. Ravel (2009). "Chapter 8. Methods for *in silico* prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data." *Methods Enzymol* 458: 181-217.
- Bailey, C. B. and C. C. Balch (1961). "Saliva secretion and its relation to feeding in cattle. 1. The composition and rate of secretion of parotid saliva in a small steer." *Br J Nutr* 15: 371-382.
- Balch, W. E. and R. S. Wolfe (1976). "New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressureized atmosphere." *Appl Environ Microbiol* 32(6): 781-791.
- Balch, W. E. and R. S. Wolfe (1979). "Transport of coenzyme M (2-mercaptoethanesulfonic acid) in *Methanobacterium ruminantium*." *J Bacteriol* 137(1): 264-273.
- Balch, W. E., G. E. Fox, *et al.* (1979). "Methanogens: reevaluation of a unique biological group." *Microbiol Rev* 43(2): 260-296.
- Balint, M., S. Domisch, *et al.* (2011). "Cryptic biodiversity loss linked to global climate change." *Nature Clim. Change* 1(6): 313-318.
- Bankevich, A., S. Nurk, *et al.* (2012). "SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing." *J Comput Biol* 19(5): 455-477.
- Bapteste, É., C. Brochier, *et al.* (2005). "Higher-level classification of the Archaea: Evolution of methanogenesis and methanogens." *Archaea* 1(5): 353-363.
- Barenkamp, S. J. (1996). "Immunization with high-molecular-weight adhesion proteins of nontypeable *Haemophilus influenzae* modifies experimental otitis media in chinchillas." *Infect Immun* 64(4): 1246-1251.
- Barenkamp, S. J. and J. W. St Geme, 3rd (1996). "Identification of a second family of high-molecular-weight adhesion proteins expressed by non-typable *Haemophilus influenzae*." *Mol Microbiol* 19(6): 1215-1223.

- Baumer, S., T. Ide, *et al.* (2000). "The F<sub>420</sub>H<sub>2</sub> dehydrogenase from *Methanosarcina mazei* is a redox-driven proton pump closely related to NADH dehydrogenases." *J Biol Chem* 275(24): 17968-17973.
- Baysse, C., S. Matthijs, *et al.* (2001). "Impact of mutations in *hemA* and *hemH* genes on pyoverdine production by *Pseudomonas fluorescens* ATCC17400." *FEMS Microbiol Lett* 205(1): 57-63.
- Beauchemin, K. A., M. Kreuzer, *et al.* (2008). "Nutritional management for enteric methane abatement: a review." *Anim Prod Sci* 48(2): 21-27.
- Beauchemin, K. A. and S. M. McGinn (2005). "Methane emissions from feedlot cattle fed barley or corn diets." *J Anim Sci* 83(3): 653-661.
- Beauchemin, K. A. and S. M. McGinn (2006). "Methane emissions from beef cattle: effects of fumaric acid, essential oil, and canola oil." *J Anim Sci* 84(6): 1489-1496.
- Beauchemin, K. A., S. M. McGinn, *et al.* (2007). "Use of condensed tannin extract from quebracho trees to reduce methane emissions from cattle." *J Anim Sci* 85(8): 1990-1996.
- Begley, M., P. D. Cotter, *et al.* (2009). "Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins." *Appl Environ Microbiol* 75(17): 5451-5460.
- Belfaiza, J., A. Fazel, *et al.* (1984). "*E. coli* aspartokinase II-homoserine dehydrogenase II polypeptide chain has a triglobular structure." *Biochem Biophys Res Commun* 123(1): 16-20.
- Bellais, S., M. Arthur, *et al.* (2006). "Asl<sub>fm</sub>, the D-aspartate ligase responsible for the addition of D-aspartic acid onto the peptidoglycan precursor of *Enterococcus faecium*." *J Biol Chem* 281(17): 11586-11594.
- Bender, D. A. (2012). *Amino Acid Metabolism*, Wiley-Blackwell.
- Benjamini, Y. (2010). "Discovering the false discovery rate." *J Royal Stat Soc: Ser B (Stat Meth)* 72(4): 405-416.
- Berg, I. A., D. Kockelkorn, *et al.* (2010). "Autotrophic carbon fixation in archaea." *Nat Rev Micro* 8(6): 447-460.

- Berge, O., T. Heulin, *et al.* (1991). "*Rahnella aquatilis*, a nitrogen-fixing enteric bacterium associated with the rhizosphere of wheat and maize." *Can J Microbiol* 37(3): 195-203.
- Berk, H. and R. K. Thauer (1997). "Function of coenzyme F<sub>420</sub>-dependent NADP reductase in methanogenic archaea containing an NADP-dependent alcohol dehydrogenase." *Arch Microbiol* 168(5): 396-402.
- Bernstein, M., F. Kepes, *et al.* (1989). "Sec59 encodes a membrane protein required for core glycosylation in *Saccharomyces cerevisiae*." *Mol Cell Biol* 9(3): 1191-1199.
- Berven, F. S., O. A. Karlsen, *et al.* (2006). "Analysing the outer membrane subproteome of *Methylococcus capsulatus* (Bath) using proteomics and novel biocomputing tools." *Arch Microbiol* 184(6): 362-377.
- Betley, J. N., M. C. Frith, *et al.* (2002). "A ubiquitous and conserved signal for RNA localization in chordates." *Curr Biol* 12(20): 1756-1761.
- Beulin, D. S., M. Yamaguchi, *et al.* (2014). "Crystal structure of PfbA, a surface adhesin of *Streptococcus pneumoniae*, provides hints into its interaction with fibronectin." *Int J Biol Macromol* 64: 168-173.
- Biemont, C. and C. Vieira (2006). "Genetics: Junk DNA as an evolutionary force." *Nature* 443(7111): 521-524.
- Bird, S. H., R. S. Hegarty, *et al.* (2008). "Persistence of defaunation effects on digestion and methane production in ewes." *Aust J Exp Agric* 48(1-2): 152-155.
- Biswas, A., J. N. Gagnon, *et al.* (2013). "CRISPRTarget: bioinformatic prediction and analysis of crRNA targets." *RNA Biol* 10(5): 817-827.
- Blanche, F., C. Robin, *et al.* (1991). "Purification, characterization, and molecular cloning of S-adenosyl-L-methionine: uroporphyrinogen III methyltransferase from *Methanobacterium ivanovii*." *J Bacteriol* 173(15): 4637-4645.
- Bray, J. R. and J. T. Curtis (1957). "An ordination of the upland forest communities of southern Wisconsin." *Ecol Monogr* 27(4): 325-349.
- Bodegom, P. M., J. C. Scholten, *et al.* (2004). "Direct inhibition of methanogenesis by ferric iron." *FEMS Microbiol Ecol* 49(2): 261-268.

- Bonacker, L. G., S. Baudner, *et al.* (1993). "Properties of the two isoenzymes of methyl-coenzyme M reductase in *Methanobacterium thermoautotrophicum*." *Eur J Biochem* 217(2): 587-595.
- Boomker, E. A. and P. Cronjé (2000). *Ruminant physiology: digestion, metabolism, growth, and reproduction*, CABI.
- Borrel, G., N. Gaci, *et al.* (2014). "Unique characteristics of the pyrrolysine system in the 7th order of methanogens: implications for the evolution of a genetic code expansion cassette." *Archaea* 2014: 374146.
- Borrel, G., H. M. Harris, *et al.* (2013). "Genome sequence of "*Candidatus* Methanomassiliicoccus intestinalis" Isoire-Mx1, a third thermoplasmatales-related methanogenic archaeon from human feces." *Genome Announc* 1(4).
- Borrel, G., H. M. Harris, *et al.* (2012). "Genome sequence of "*Candidatus* Methanomethylophilus alvus" Mx1201, a methanogenic archaeon from the human gut belonging to a seventh order of methanogens." *J Bacteriol* 194(24): 6944-6945.
- Borrel, G., P. W. O'Toole, *et al.* (2013). "Phylogenomic data support a seventh order of Methylophilic methanogens and provide insights into the evolution of Methanogenesis." *Genome Biol Evol* 5(10): 1769-1780.
- Borrel, G., N. Parisot, *et al.* (2014). "Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine." *BMC Genomics* 15(1): 679.
- Bose, A., M. A. Pritchett, *et al.* (2006). "Differential regulation of the three methanol methyltransferase isozymes in *Methanosarcina acetivorans* C2A." *J Bacteriol* 188(20): 7274-7283.
- Bose, A., M. A. Pritchett, *et al.* (2008). "Genetic analysis of the methanol- and methylamine-specific methyltransferase 2 genes of *Methanosarcina acetivorans* C2A." *J Bacteriol* 190(11): 4017-4026.
- Boudsocq, F., S. Iwai, *et al.* (2001). "*Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta." *Nucleic Acids Res* 29(22): 4607-4616.



- Bower, S., J. B. Perkins, *et al.* (1996). "Cloning, sequencing, and characterization of the *Bacillus subtilis* biotin biosynthetic operon." *J Bacteriol* 178(14): 4122-4130.
- Bray, J. R. and J. T. Curtis (1957). "An ordination of the upland forest communities of southern Wisconsin." *Ecol Monogr* 27(4): 325-349.
- Bremer, H. and P. P. Dennis (2008). "Modulation of chemical composition and other parameters of the cell at different exponential growth rates." *EcoSal Plus* 3(1).
- Brodsky, L., A. Moussaieff, *et al.* (2010). "Evaluation of peak picking quality in LC-MS metabolomics data." *Anal Chem* 82(22): 9177-9187.
- Brosius, J., T. J. Dull, *et al.* (1981). "Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*." *J Mol Biol* 148(2).
- Brügger, K., P. Redder, *et al.* (2002). "Mobile elements in archaeal genomes." *FEMS Microbiol Lett* 206(2).
- Brune, A. (1998). "Termite guts: the world's smallest bioreactors." *Trends Biotechnol* 16(1): 16-21.
- Bryant, M. P. (1965). *Rumen methanogenic bacteria. Physiology of digestion of the ruminant.* R. W. Dougherty, Butterworths Publishing Inc.
- Bryant, M. P. and N. Small (1956). "Characteristics of two new genera of anaerobic curved rods isolated from the rumen of cattle." *J Bacteriol* 72(1): 22-26.
- Bryant, M. P., L. L. Campbell, *et al.* (1977). "Growth of *Desulfovibrio* in lactate or ethanol media low in sulfate in association with H<sub>2</sub>-utilizing methanogenic bacteria." *Appl Environ Microbiol* 33(5): 1162-1169.
- Bryant, M., S. Tzeng, *et al.* (1971). "Nutrient requirements of methanogenic bacteria." *Urbana* 101: 61801.
- Buan, N. R. and W. W. Metcalf (2010). "Methanogenesis by *Methanosarcina acetivorans* involves two structurally and functionally distinct classes of heterodisulfide reductase." *Mol Microbiol* 75(4): 843-853.

- Buckel, W. and R. K. Thauer (2013). "Energy conservation via electron bifurcating ferredoxin reduction and proton/Na<sup>+</sup> translocating ferredoxin oxidation." *Biochim Biophys Acta* 1827(2): 94-113.
- Buddle, B. M., M. Denis, *et al.* (2011). "Strategies to reduce methane emissions from farmed ruminants grazing on pasture." *Vet J* 188(1): 11-17.
- Buhk, H. J., B. Behrens, *et al.* (1984). "Restriction and modification in *Bacillus subtilis*: nucleotide sequence, functional organization and product of the DNA methyltransferase gene of bacteriophage SPR." *Gene* 29(1-2): 51-61.
- Bullard, J. H., E. Purdom, *et al.* (2010). "Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments." *BMC Bioinformatics* 11: 94.
- Bult, C. J., O. White, *et al.* (1996). "Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*." *Science* 273(5278): 1058-1073.
- Burke, S. A. and J. A. Krzycki (1997). "Reconstitution of Monomethylamine:Coenzyme M methyl transfer with a corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*." *J Biol Chem* 272(26): 16570-16577.
- Burrows, M. and D. J. Wheeler (1994). "A block-sorting lossless data compression algorithm."
- Bygrave, F. L. and R. M. Dawson (1976). "Phosphatidylcholine biosynthesis and choline transport in the anaerobic protozoon *Entodinium caudatum*." *Biochem J* 160(3): 481-490.
- Cabanes, D., P. Dehoux, *et al.* (2002). "Surface proteins and the pathogenic potential of *Listeria monocytogenes*." *Trends Microbiol* 10(5): 238-245.
- Caboche, S., M. Pupin, *et al.* (2008). "NORINE: a database of nonribosomal peptides." *Nucleic Acids Res* 36(Database issue): D326-331.
- Canchaya, C., C. Proux, *et al.* (2003). "Prophage genomics." *Microbiol Mol Biol Rev* 67(2): 238-276.
- Candlish, E., T. J. Devlin, *et al.* (1970). "Tryptophan utilization by rumen microorganisms *in vitro*." *Can J Anim Sci* 50(2): 331-335.
- Carulla, J., M. Kreuzer, *et al.* (2005). "Supplementation of *Acacia mearnsii* tannins decreases methanogenesis and urinary nitrogen in forage-fed sheep." *Crop Pasture Sci* 56(9): 961-970.

- Carver, T., N. Thomson, *et al.* (2009). "DNAPlotter: circular and linear interactive genome visualization." *Bioinformatics* (Oxford, England) 25(1): 119-120.
- Casali, N. and A. Preston (2003). *E. coli* plasmid vectors: Methods and applications, Springer Science & Business Media.
- Caspi, R., T. Altman, *et al.* (2014). "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases." *Nucleic Acids Res* 42(Database issue): D459-471.
- Chagan, I., M. Tokura, *et al.* (1999). "Detection of methanogenic archaea associated with rumen ciliate protozoa." *J Gen Appl Microbiol* 45(6): 305-308.
- Chant, J. and P. Dennis (1986). "Archaeobacteria: transcription and processing of ribosomal RNA sequences in *Halobacterium cutirubrum*." *EMBO J* 5(5): 1091-1097.
- Chaudhary, P. P., S. K. Sirohi, *et al.* (2011). "Methyl coenzyme M reductase (*mcrA*) gene based phylogenetic analysis of methanogens population in Murrah buffaloes (*Bubalus bubalis*)." *J Microbiol* 49(4): 558-561.
- Cheeseman, P., A. Toms-Wood, *et al.* (1972). "Isolation and properties of a fluorescent compound, Factor<sub>420</sub>, from *Methanobacterium* strain M.o.H." *J Bacteriol* 112(1): 527-531.
- Chen, H., N. Huang, *et al.* (2006). "SubLoc: a server/client suite for protein subcellular location based on SOAP." *Bioinformatics* 22(3): 376-377.
- Chen, S., L. Wang, *et al.* (2010). "Twenty years hunting for sulfur in DNA." *Protein Cell* 1(1): 14-21.
- Christensen, S. K. and K. Gerdes (2003). "RelE toxins from bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA." *Mol Microbiol* 48(5): 1389-1400.
- Church, J. A. and N. J. White (2011). "Sea-level rise from the late 19th to the early 21st century." *Surv Geophys* 32(4): 585-602.
- Clark, H., F. Kelliher, *et al.* (2011). "Reducing CH<sub>4</sub> emissions from grazing ruminants in New Zealand: challenges and opportunities." *Asian Aust J Anim Sci* 24(2): 295-302.

- Clouet d'Orval, B., M. L. Bortolin, *et al.* (2001). "Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNA<sup>Trp</sup> intron guides the formation of two ribose-methylated nucleosides in the mature tRNA<sup>Trp</sup>." *Nucleic Acids Res* 29(22): 4518-4529.
- Conrad, R., C. Erkel, *et al.* (2006). "Rice Cluster I methanogens, an important group of Archaea producing greenhouse gas in soil." *Curr Opin Biotechnol* 17(3): 262-267.
- Corcelli, A. (2009). "The cardiolipin analogues of Archaea." *Biochim Biophys Acta* 1788(10): 2101-2106.
- Craciun, S. and E. P. Balskus (2012). "Microbial conversion of choline to trimethylamine requires a glycol radical enzyme." *Proc Natl Acad Sci U S A* 109(52): 21307-21312.
- Creevey, C. J., W. J. Kelly, *et al.* (2014). "Determining the culturability of the rumen bacterial microbiome." *Microb Biotechnol* 7(5): 467-479.
- Cubasch, U., D. Wuebbles, *et al.* (2013). Introduction. *Climate Change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. T. F. Stocker, D. Qin, G.-K. Plattner *et al.* Cambridge, United Kingdom and New York, NY, USA, Cambridge University Press: 119–158.
- Cunin, R., N. Glansdorff, *et al.* (1986). "Biosynthesis and metabolism of arginine in bacteria." *Microbiol Rev* 50(3): 314-352.
- Currie, D. H., B. Raman, *et al.* (2015). "Genome-scale resources for *Thermoanaerobacterium saccharolyticum*." *BMC Syst Biol* 9: 30.
- D'Souza, G., S. Waschina, *et al.* (2014). "Less is more: selective advantages can explain the prevalent loss of biosynthetic genes in bacteria." *Evolution* 68(9): 2559-2570.
- Dabard, J., C. Bridonneau, *et al.* (2001). "Ruminococcin A, a new lantibiotic produced by a *Ruminococcus gnavus* strain isolated from human feces." *Appl Environ Microbiol* 67(9): 4111-4118.
- De Poorter, L. M., W. G. Geerts, *et al.* (2003). "Bioenergetics of the formyl-methanofuran dehydrogenase and heterodisulfide reductase reactions in *Methanothermobacter thermautotrophicus*." *Eur J Biochem* 270(1): 66-75.

De Vrije, T. and P. A. M. Claassen (2003). Dark hydrogen fermentations. Bio-methane & Bio-hydrogen. J. H. Reith, R. H. Wijffels and H. Barten. Netherlands, Dutch Biological Hydrogen Foundation: 103-123.

Dehority, B. A. (1969). "Pectin-fermenting bacteria isolated from the bovine rumen." J Bacteriol 99(1): 189-196.

Dennis, P. P. (1997). "Ancient ciphers: translation in Archaea." Cell 89(7): 1007-1010.

Dejean, A., P. Sonigo, *et al.* (1984). "Specific hepatitis B virus integration in hepatocellular carcinoma DNA through a viral 11-base-pair direct repeat." Proc Natl Acad Sci U S A 81(17): 5350-5354.

Dejean, S., I. Gonzalez, *et al.* (2011). "Package 'mixOmics'." Image 9: 1.

Delcher, A. L., D. Harmon, *et al.* (1999). "Improved microbial gene identification with GLIMMER." Nucleic Acids Res 27(23): 4636-4641.

Delcher, A. L., S. L. Salzberg, *et al.* (2003). "Using MUMmer to identify similar regions in large sequence sets." Curr Protoc Bioinformatics Chapter 10: Unit 10 13.

Dennis, P. P. (1997). "Ancient ciphers: translation in Archaea." Cell 89(7): 1007-1010.

Deppenmeier, U. (2002). "The unique biochemistry of methanogenesis." Prog Nucleic Acid Res Mol Biol 73: 223-283.

Deppenmeier, U. (2004). "The membrane-bound electron transport system of *Methanosarcina* species." J Bioenerg Biomembr 36(1): 55-64.

Deppenmeier, U., L. Tanja, *et al.* (1999). "Novel reactions involved in energy conservation by methanogenic archaea." FEBS Lett 457(3): 291-297.

Dermoumi, H. L. and R. A. Ansorg (2001). "Isolation and antimicrobial susceptibility testing of fecal strains of the archaeon *Methanobrevibacter smithii*." Chemotherapy 47(3): 177-183.

Devenish, S. R. A., J. W. Blunt, *et al.* (2010). "NMR studies uncover alternate substrates for dihydrodipicolinate synthase and suggest that dihydrodipicolinate reductase is also a dehydratase." J Med Chem 53(12): 4808-4812.

- Diekert, G., U. Konheiser, *et al.* (1981). "Nickel requirement and factor F<sub>430</sub> content of methanogenic bacteria." *J Bacteriol* 148(2): 459-464.
- Ding, S. Y., M. T. Rincon, *et al.* (2001). "Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*." *J Bacteriol* 183(6): 1945-1953.
- Doddema, H. J. and G. D. Vogels (1978). "Improved identification of methanogenic bacteria by fluorescence microscopy." *Appl Environ Microbiol* 36(5): 752-754.
- Dot, M., J. T. Roehr, *et al.* (2012). "FLEXBAR-Flexible barcode and adapter processing for next-generation sequencing platforms." *Biology (Basel)* 1(3): 895-905.
- Dridi, B., M. L. Fardeau, *et al.* (2012). "*Methanomassiliicoccus luminyensis* gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces." *Int J Syst Evol Microbiol* 62(Pt 8): 1902-1907.
- Dridi, B., M. Henry, *et al.* (2012). "Age-related prevalence of *Methanomassiliicoccus luminyensis* in the human gut microbiome." *APMIS* 120(10): 773-777.
- Dumora, C., A. M. Lacoste, *et al.* (1983). "Purification and properties of 2-aminoethylphosphonate:pyruvate aminotransferase from *Pseudomonas aeruginosa*." *Eur J Biochem* 133(1): 119-125.
- Dykhuizen, D. (1978). "Selection for tryptophan auxotrophs of *Escherichia coli* in glucose-limited chemostats as a test of the energy conservation hypothesis of evolution." *Evolution*: 125-150.
- Eckburg, P. B., E. M. Bik, *et al.* (2005). "Diversity of the human intestinal microbial flora." *Science* 308(5728): 1635-1638.
- Eckard, R. J., C. Grainger, *et al.* (2010). "Options for the abatement of methane and nitrous oxide from ruminant production: A review." *Livest Sci* 130(1-3): 47-56.
- Eddy, S. R. (1998). "Profile hidden Markov models." *Bioinformatics* 14(9): 755-763.
- Eddy, S. R. and R. Durbin (1994). "RNA sequence analysis using covariance models." *Nucleic Acids Res* 22(11): 2079-2088.

- Egert, M., B. Wagner, *et al.* (2003). "Microbial community structure in midgut and hindgut of the humus-feeding larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae)." *Appl Environ Microbiol* 69(11): 6659-6668.
- Ehrlich, M., K. F. Norris, *et al.* (1986). "DNA cytosine methylation and heat-induced deamination." *Biosci Rep* 6(4): 387-393.
- Eichler, J. and R. Moll (2001). "The signal recognition particle of Archaea." *Trends Microbiol* 9(3): 130-136.
- Eisen, J. A., J. F. Heidelberg, *et al.* (2000). "Evidence for symmetric chromosomal inversions around the replication origin in bacteria." *Genome Biol* 1(6): RESEARCH0011.
- Ekiel, I., G. D. Sprott, *et al.* (1985). "Acetate and CO<sub>2</sub> assimilation by *Methanothrix concilii*." *J Bacteriol* 162(3): 905-908.
- Ellefson, W. L. and R. S. Wolfe (1981). "Component C of the methylreductase system of *Methanobacterium*." *J Biol Chem* 256(9): 4259-4262.
- Ellefson, W. L., W. B. Whitman, *et al.* (1982). "Nickel-containing factor F<sub>430</sub>: chromophore of the methylreductase of *Methanobacterium*." *Proc Natl Acad Sci U S A* 79(12): 3707-3710.
- Ellen, A. F., B. Zolghadr, *et al.* (2010). "Shaping the archaeal cell envelope." *Archaea* 2010: 608243.
- Ellen, A. F., S. V. Albers, *et al.* (2009). "Proteomic analysis of secreted membrane vesicles of archaeal *Sulfolobus* species reveals the presence of endosome sorting complex components." *Extremophiles* 13(1): 67-79.
- Ellen, A. F., S. V. Albers, *et al.* (2010). "Comparative study of the extracellular proteome of *Sulfolobus* species reveals limited secretion." *Extremophiles* 14(1): 87-98.
- Ermler, U., W. Grabarse, *et al.* (1997). "Crystal structure of methyl-coenzyme M reductase: the key enzyme of biological methane formation." *Science* 278(5342): 1457-1462.
- Evans, P. N., D. H. Parks, *et al.* (2015). "Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics." *Science* 350(6259): 434-438.

- Ezaki, S., N. Maeda, *et al.* (1999). "Presence of a structurally novel type ribulose-bisphosphate carboxylase/oxygenase in the hyperthermophilic archaeon, *Pyrococcus kodakaraensis* KOD1." *J Biol Chem* 274(8): 5078-5082.
- Fabry, V. J., B. A. Seibel, *et al.* (2008). "Impacts of ocean acidification on marine fauna and ecosystem processes." *ICES J Marine Sci* 65(3): 414-432.
- Fay, M. and M. M. Fay (2010). "Package 'perm'."
- Felsenstein, J. (1985). "Confidence limits on phylogenies: An approach using the Bootstrap." *Evolution* 39(4): 783-791.
- Fenchel, T. and B. J. Finlay (1991). "The biology of free-living anaerobic ciliates." *Europ J Protistol* 26(3): 201 - 215.
- Ferguson, D. J., Jr., N. Gorlatova, *et al.* (2000). "Reconstitution of dimethylamine:coenzyme M methyl transfer with a discrete corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*." *J Biol Chem* 275(37): 29053-29060.
- Ferguson, D. J., Jr. and J. A. Krzycki (1997). "Reconstitution of trimethylamine-dependent coenzyme M methylation with the trimethylamine corrinoid protein and the isozymes of methyltransferase II from *Methanosarcina barkeri*." *J Bacteriol* 179(3): 846-852.
- Ferrari, A., T. Brusa, *et al.* (1994). "Isolation and characterization of *Methanobrevibacter oralis* sp. nov." *Curr Microbiol* 29(1): 7-12.
- Fiebig, K. and B. Friedrich (1989). "Purification of the F<sub>420</sub>-reducing hydrogenase from *Methanosarcina barkeri* (strain Fusaro)." *Eur J Biochem* 184(1): 79-88.
- Fiebig, K. and G. Gottschalk (1983). "Methanogenesis from choline by a coculture of *Desulfovibrio* sp. and *Methanosarcina barkeri*." *Appl Environ Microbiol* 45(1): 161-168.
- Fimereli, D. K., K. D. Tsirigos, *et al.* (2012). CW-PRED: a HMM-based method for the classification of cell wall-anchored proteins of Gram-positive bacteria. *Artificial Intelligence: Theories and Applications*, Springer: 285-290.
- Finlay, B. J., G. Esteban, *et al.* (1994). "Some rumen ciliates have endosymbiotic methanogens." *FEMS Microbiol Lett* 117(2): 157-161.



Finn, R. D., P. Coghill, *et al.* (2016). "The Pfam protein families database: towards a more sustainable future." *Nucleic Acids Res* 44(D1): D279-285.

Fischer, M., W. Römisch, *et al.* (2002). "Biosynthesis of riboflavin in archaea studies on the mechanism of 3, 4-dihydroxy-2-butanone-4-phosphate synthase of *Methanococcus jannaschii*." *J Biol Chem* 277(44): 41410-41416.

Fischer, M., A. K. Schott, *et al.* (2004). "Evolution of vitamin B2 biosynthesis. A novel class of riboflavin synthase in Archaea." *J Mol Biol* 343(1): 267-278.

Fischer, R. and R. Jensen (1987). "Prephenate dehydratase (monofunctional)." *Methods Enzymol* 142: 507-512.

Fischer, R. J., J. Helms, *et al.* (1993). "Cloning, sequencing, and molecular analysis of the sol operon of *Clostridium acetobutylicum*, a chromosomal locus involved in solventogenesis." *J Bacteriol* 175(21): 6959-6969.

Flannigan, K. A., S. H. Hennigan, *et al.* (1990). "Purine biosynthesis in *Escherichia coli* K12: structure and DNA sequence studies of the *purHD* locus." *Mol Microbiol* 4(3): 381-392.

Flint, D. H., J. F. Tuminello, *et al.* (1996). "Studies on the synthesis of the Fe-S cluster of dihydroxy-acid dehydratase in *Escherichia coli* crude extract. Isolation of *O*-acetylserine sulfhydrylases A and B and  $\beta$ -cystathionase based on their ability to mobilize sulfur from cysteine and to participate in Fe-S cluster synthesis." *J Biol Chem* 271(27): 16053-16067.

Foley, P. A., D. A. Kenny, *et al.* (2009). "Effect of DL-malic acid supplementation on feed intake, methane emission, and rumen fermentation in beef cattle." *J Anim Sci* 87(3): 1048-1057.

Fournier, G. P., A. A. Dick, *et al.* (2011). "Evolution of the Archaea: emerging views on origins and phylogeny." *Res Microbiol* 162(1): 92-98.

Fournier, G. P. and J. P. Gogarten (2008). "Evolution of acetoclastic methanogenesis in *Methanosarcina* via horizontal gene transfer from cellulolytic *Clostridia*." *J Bacteriol* 190(3): 1124-1127.

Fouts, D. E., E. F. Mongodin, *et al.* (2007). Genome sequence of *Campylobacter curvus* 525.92 isolated from human feces, EMBL/GenBank/DDBJ databases.

- Franzosa, E. A., X. C. Morgan, *et al.* (2014). "Relating the metatranscriptome and metagenome of the human gut." *Proc Natl Acad Sci U S A* 111(22): E2329-2338.
- Fraser, H. I., M. Kvaratskhelia, *et al.* (1999). "The two analogous phosphoglycerate mutases of *Escherichia coli*." *FEBS Lett* 455(3): 344-348.
- Fricke, W. F., H. Seedorf, *et al.* (2006). "The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H<sub>2</sub> for methane formation and ATP synthesis." *J Bacteriol* 188(2): 642-658.
- Friedrich, M. W. (2005). Methyl - coenzyme M reductase genes: unique functional markers for methanogenic and anaerobic methane - oxidizing Archaea. *Methods Enzymol.* R. L. Jared, Academic Press. Volume 397: 428-442.
- Frost, L. S., R. Leplae, *et al.* (2005). "Mobile genetic elements: the agents of open source evolution." *Nat Rev Microbiol* 3(9): 722-732.
- Fukuda, W., N. Morimoto, *et al.* (2008). "Agmatine is essential for the cell growth of *Thermococcus kodakaraensis*." *FEMS Microbiol Lett* 287(1): 113-120.
- Fuqua, C. (2010). "Passing the baton between laps: adhesion and cohesion in *Pseudomonas putida* biofilms." *Mol Microbiol* 77(3): 533-536.
- Furukawa, N., A. Miyanaga, *et al.* (2014). "Diverse allosteric and catalytic functions of tetrameric D-lactate dehydrogenases from three Gram-negative bacteria." *AMB Express* 4: 76.
- Gao, B. and R. S. Gupta (2007). "Phylogenomic analysis of proteins that are distinctive of Archaea and its main subgroups and the origin of methanogenesis." *BMC Genomics* 8.
- Garrity, G. M., T. G. Lilburn, *et al.* (2007). "Part 1-The Archaea: Phyla Crenarchaeota and Euryarchaeota." *TOBA* 7(7): 6-31.
- Gaston, M. A., R. Jiang, *et al.* (2011). "Functional context, biosynthesis, and genetic encoding of pyrrolysine." *Curr Opin Microbiol* 14(3): 342-349.
- Gaston, M. A., L. Zhang, *et al.* (2011). "The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine." *Nature* 471(7340): 647-650.

- Geerkens, C. H., R. M. Schweiggert, *et al.* (2013). "Influence of apple and citrus pectins, processed mango peels, a phenolic mango peel extract, and gallic acid as potential feed supplements on *in vitro* total gas production and rumen methanogenesis." *J Agric Food Chem* 61(24): 5727-5737.
- Gerdes, K., S. K. Christensen, *et al.* (2005). "Prokaryotic toxin-antitoxin stress response loci." *Nat Rev Microbiol* 3(5): 371-382.
- Ghali, M. B., P. T. Scott, *et al.* (2004). "Characterization of *Streptococcus bovis* from the rumen of the dromedary camel and Rusa deer." *Lett Appl Microbiol* 39(4): 341-346.
- Gilbert, C. and R. Cordaux (2013). "Horizontal transfer and evolution of prokaryote transposable elements in eukaryotes." *Genome Biol Evol* 5(5): 822-832.
- Gish, W. and D. J. States (1993). "Identification of protein coding regions by database similarity search." *Nat Genet* 3(3): 266-272.
- Goel, G. and H. P. Makkar (2012). "Methane mitigation from ruminants using tannins and saponins." *Trop Anim Health Prod* 44(4): 729-739.
- Goldberg, T., M. Hecht, *et al.* (2014). "LocTree3 prediction of localization." *Nucleic Acids Res* 42(Web Server issue): W350-355.
- Gomez-Alarcon, R. A., C. O'Dowd, *et al.* (1982). "1,4-Naphthoquinone and other nutrient requirements of *Succinivibrio dextrinosolvens*." *Appl Environ Microbiol* 44(2): 346-350.
- Gorlas, A., C. Robert, *et al.* (2012). "Complete genome sequence of *Methanomassiliicoccus luminyensis*, the largest genome of a human-associated Archaea species." *J Bacteriol* 194(17): 4745.
- Gottschalk, G. and R. K. Thauer (2001). "The Na<sup>+</sup>-translocating methyltransferase complex from methanogenic archaea." *Biochim Biophys Acta* 1505(1): 28-36.
- Gower, J. C. and P. Legendre (1986). "Metric and Euclidean properties of dissimilarity coefficients." *J Classification* 3(1): 5-48.
- Grabarse, W., F. Mählert, *et al.* (2000). "Comparison of three methyl-coenzyme M reductases from phylogenetically distant organisms: unusual amino acid modification, conservation and adaptation." *J Mol Biol* 303(2): 329-344.

- Graber, J. R. and J. A. Breznak (2005). "Folate cross-feeding supports symbiotic homoacetogenic spirochetes." *Appl Environ Microbiol* 71(4): 1883-1889.
- Grada, A. and K. Weinbrecht (2013). "Next-generation sequencing: methodology and application." *J Invest Dermatol* 133(8): e11.
- Graham, D. E., S. M. Taylor, *et al.* (2009). "Convergent evolution of coenzyme M biosynthesis in the *Methanosarcinales*: cysteate synthase evolved from an ancestral threonine synthase." *Biochem J* 424(3): 467-478.
- Green, S. M., T. Malik, *et al.* (1996). "The *purB* gene of *Escherichia coli* K-12 is located in an operon." *Microbiology* 142 ( Pt 11): 3219-3230.
- Griffin, M., R. Casadio, *et al.* (2002). "Transglutaminases: nature's biological glues." *Biochem J* 368(Pt 2): 377-396.
- Griffiths-Jones, S., S. Moxon, *et al.* (2005). "Rfam: annotating non-coding RNAs in complete genomes." *Nucleic Acids Res* 33(Database issue): D121-124.
- Grissa, I., G. Vergnaud, *et al.* (2007). "CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats." *Nucleic Acids Res* 35(Web Server issue): W52-57.
- Grochowski, L. L., H. Xu, *et al.* (2006). "*Methanocaldococcus jannaschii* uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate." *J Bacteriol* 188(9): 3192-3198.
- Großkopf, R., S. Stubner, *et al.* (1998). "Novel euryarchaeotal lineages detected on rice roots and in the anoxic bulk soil of flooded rice microcosms." *Appl Environ Microbiol* 64(12): 4983-4989.
- Gruer, M. J. and J. R. Guest (1994). "Two genetically-distinct and differentially-regulated aconitases (AcnA and AcnB) in *Escherichia coli*." *Microbiology* 140 ( Pt 10): 2531-2541.
- Guan, H., K. M. Wittenberg, *et al.* (2006). "Efficacy of ionophores in cattle diets for mitigation of enteric methane." *J Anim Sci* 84(7): 1896-1906.

- Gunsalus, R. P. and R. S. Wolfe (1980). "Methyl coenzyme M reductase from *Methanobacterium thermoautotrophicum*. Resolution and properties of the components." J Biol Chem 255(5): 1891-1895.
- Guo, Y. Q., J. X. Liu, *et al.* (2008). "Effect of tea saponin on methanogenesis, microbial community structure and expression of *mcrA* gene, in cultures of rumen micro-organisms." Lett Appl Microbiol 47(5): 421-426.
- Gupta, A. and U. B. Chaudhary (2010). "Isolation and characterization of *Methanobrevibacter smithii* GMS-01 from rumen of goats." Indian Vet J 87(10): 1009-1012.
- Guss, A. M., G. Kulkarni, *et al.* (2009). "Differences in hydrogenase gene expression between *Methanosarcina acetivorans* and *Methanosarcina barkeri*." J Bacteriol 191(8): 2826-2833.
- Haase, I., S. Mortl, *et al.* (2003). "Biosynthesis of riboflavin in archaea. 6,7-dimethyl-8-ribityllumazine synthase of *Methanococcus jannaschii*." Eur J Biochem 270(5): 1025-1032.
- Haft, D. H., J. D. Selengut, *et al.* (2003). "The TIGRFAMs database of protein families." Nucleic Acids Res 31(1): 371-373.
- Haines, T. H. and N. A. Dencher (2002). "Cardiolipin: a proton trap for oxidative phosphorylation." FEBS Lett 528(1-3): 35-39.
- Hamby, S. E., N. S. Thomas, *et al.* (2011). "A meta-analysis of single base-pair substitutions in translational termination codons ('nonstop' mutations) that cause human inherited disease." Hum Genomics 5(4): 241-264.
- Hansen, T., D. Wendorff, *et al.* (2004). "Bifunctional phosphoglucose/phosphomannose isomerases from the Archaea *Aeropyrum pernix* and *Thermoplasma acidophilum* constitute a novel enzyme family within the phosphoglucose isomerase superfamily." J Biol Chem 279(3): 2262-2272.
- Hao, B., W. Gong, *et al.* (2002). "A new UAG-encoded residue in the structure of a methanogen methyltransferase." Science 296(5572): 1462-1466.
- Hartmann, D. L., A. M. G. Klein Tank, *et al.* (2013). Observations: Atmosphere and Surface. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. T. F. Stocker, D.

Qin, G.-K. Plattner *et al.* Cambridge, United Kingdom and New York, NY, USA, Cambridge University Press: 159–254.

Harms, U., D. S. Weiss, *et al.* (1995). "The energy conserving  $N^5$ -methyltetrahydromethanopterin:coenzyme M methyltransferase complex from *Methanobacterium thermoautotrophicum* is composed of eight different subunits." *Eur J Biochem* 228(3): 640-648.

Haruyama, K., T. Nakai, *et al.* (2001). "Structures of *Escherichia coli* histidinol-phosphate aminotransferase and its complexes with histidinol-phosphate and *N*-(5'-phosphopyridoxyl)-L-glutamate: double substrate recognition of the enzyme." *Biochem* 40(15): 4633-4644.

Hayes, F. (2003). "Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest." *Science* 301(5639): 1496-1499.

Hedderich, R. (2004). "Energy-converting [NiFe] hydrogenases from archaea and extremophiles: ancestors of complex I." *J Bioenerg Biomembr* 36(1): 65-75.

Hedderich, R., A. Berkessel, *et al.* (1990). "Purification and properties of heterodisulfide reductase from *Methanobacterium thermoautotrophicum* (strain Marburg)." *Eur J Biochem* 193(1): 255-261.

Hedderich, R., N. Hamann, *et al.* (2005). "Heterodisulfide reductase from methanogenic archaea: a new catalytic role for an iron-sulfur cluster." *Biol Chem* 386(10): 961-970.

Heiden, S., R. Hedderich, *et al.* (1994). "Purification of a two-subunit cytochrome-*b*-containing heterodisulfide reductase from methanol-grown *Methanosarcina barkeri*." *Eur J Biochem* 221(2): 855-861.

Helgadóttir, S., G. Rosas-Sandoval, *et al.* (2007). "Biosynthesis of phosphoserine in the *Methanococcales*." *J Bacteriol* 189(2): 575-582.

Helling, R. B. (1998). "Pathway choice in glutamate synthesis in *Escherichia coli*." *J Bacteriol* 180(17): 4571-4575.

Hemmi, H., K. Shibuya, *et al.* (2004). "(S)-2, 3-Di-*O*-geranylgeranylglyceryl phosphate synthase from the thermoacidophilic archaeon *Sulfolobus solfataricus* molecular cloning and characterization of a membrane-intrinsic prenyl transferase involved in the biosynthesis of archaeal ether-linked membrane lipids." *J Biol Chem* 279(48): 50197-50203.

Henikoff, S. and J. G. Henikoff (1992). "Amino acid substitution matrices from protein blocks." *Proc Natl Acad Sci U S A* 89(22): 10915-10919.

Henderson, G., F. Cox, *et al.* (2015). "Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range." *Sci Rep* 5: 14567.

Hendrickson, E. L. and J. A. Leigh (2008). "Roles of coenzyme F<sub>420</sub>-reducing hydrogenases and hydrogen- and F<sub>420</sub>-dependent methylenetetrahydromethanopterin dehydrogenases in reduction of F<sub>420</sub> and production of hydrogen during methanogenesis." *J Bacteriol* 190(14): 4818-4821.

Herring, S., A. Ambrogelly, *et al.* (2007). "Recognition of pyrrolysine tRNA by the *Desulfitobacterium hafniense* pyrrolysyl-tRNA synthetase." *Nucleic Acids Res* 35(4): 1270-1278.

Hertz, A. F. and A. Newton (1913). "The normal movements of the colon in man." *J Physiol* 47(1-2): 57-65.

Hino, T., K. Takeshi, *et al.* (1993). "Effects of aibellin, a novel peptide antibiotic, on rumen fermentation *in vitro*." *J Dairy Sci* 76(8): 2213-2221.

Hippe, H., D. Caspari, *et al.* (1979). "Utilization of trimethylamine and other *N*-methyl compounds for growth and methane formation by *Methanosarcina barkeri*." *Proc Natl Acad Sci U S A* 76(1): 494-498.

Hirabayashi, T., T. J. Larson, *et al.* (1976). "Membrane-associated phosphatidylglycerophosphate synthetase from *Escherichia coli*: purification by substrate affinity chromatography on cytidine 5'-diphospho-1,2-diacyl-*sn*-glycerol sepharose." *Biochem* 15(24): 5205-5211.

Ho, C. L. and K. Saito (2001). "Molecular biology of the plastidic phosphorylated serine biosynthetic pathway in *Arabidopsis thaliana*." *Amino Acids* 20(3): 243-259.

Hobson, P. N. and C. S. Stewart (1997). *The rumen microbial ecosystem*. London, Blackie Academic & Professional.

Hoch, J. A., R. Losick, *et al.* (1993). *Bacillus subtilis* and other Gram-positive bacteria : biochemistry, physiology, and molecular genetics. Washington, D.C, American Society for Microbiology.

Hollander, M., D. A. Wolfe, *et al.* (2013). Nonparametric statistical methods, John Wiley & Sons.

Honzatko, R. B. and H. J. Fromm (1999). "Structure-function studies of adenylosuccinate synthetase from *Escherichia coli*." Arch Biochem Biophys 370(1): 1-8.

Hor, L., R. C. J. Dobson, *et al.* (2010). "Crystallization and preliminary X-ray diffraction analysis of diaminopimelate epimerase from *Escherichia coli*." Struct Biol and Crystal Commun 66(1): 37-40.

Horz, H. P., I. Seyfarth, *et al.* (2012). "McrA and 16S rRNA gene analysis suggests a novel lineage of *Archaea* phylogenetically affiliated with *Thermoplasmatales* in human subgingival plaque." Anaerobe 18(3): 373-377.

Hoskins, A. A., R. Anand, *et al.* (2004). "The formylglycinamide ribonucleotide amidotransferase complex from *Bacillus subtilis*: metabolite-mediated complex formation." Biochemistry 43(32): 10314-10327.

Hottes, A. K., P. L. Freddolino, *et al.* (2013). "Bacterial adaptation through loss of function." PLoS Genet 9(7): e1003617.

Hove-Jensen, B., K. W. Harlow, *et al.* (1986). "Phosphoribosylpyrophosphate synthetase of *Escherichia coli*. Properties of the purified enzyme and primary structure of the *prs* gene." J Biol Chem 261(15): 6765-6771.

Højsgaard, S., M. S. Højsgaard, *et al.* (2006). "The doBy package." The Newsletter of the R Project 6: 1.

Hsu, S. T., E. Breukink, *et al.* (2004). "The nisin-lipid II complex reveals a pyrophosphate cage that provides a blueprint for novel antibiotics." Nat Struct Mol Biol 11(10): 963-967.

Huang, L. N., H. Zhou, *et al.* (2002). "Diversity and structure of the archaeal community in the leachate of a full-scale recirculating landfill as examined by direct 16S rRNA gene sequence retrieval." FEMS Microbiol Lett 214(2): 235-240.



- Huang, L. N., Y. Q. Chen, *et al.* (2003). "Characterization of methanogenic Archaea in the leachate of a closed municipal solid waste landfill." *FEMS Microbiol Ecol* 46(2): 171-177.
- Huang, X. D., H. Y. Tan, *et al.* (2012). "Comparison of methanogen diversity of yak (*Bos grunniens*) and cattle (*Bos taurus*) from the Qinghai-Tibetan plateau, China." *BMC Microbiol* 12: 237.
- Huisingh, J., J. J. McNeill, *et al.* (1974). "Sulfate reduction by a *Desulfovibrio* species isolated from sheep rumen." *Appl Microbiol* 28(3): 489-497.
- Hungate, R. E. (1966). *The rumen and its microbes*. Michigan, Academic Press.
- Hungate, R. E. (1967). "Hydrogen as an intermediate in the rumen fermentation." *Arch Mikrobiol* 59(1-3): 158-164.
- Huntington, G. B. (1997). "Starch utilization by ruminants: from basics to the bunk." *J Anim Sci* 75(3): 852-867.
- Huss, R. J. and M. Glaser (1983). "Identification and purification of an adenylate kinase-associated protein that influences the thermolability of adenylate kinase from a temperature-sensitive *adk* mutant of *Escherichia coli*." *J Biol Chem* 258(21): 13370-13376.
- Iannotti, E. L., D. Kafkewitz, *et al.* (1973). "Glucose fermentation products in *Ruminococcus albus* grown in continuous culture with *Vibrio succinogenes*: changes caused by interspecies transfer of H<sub>2</sub>." *J Bacteriol* 114(3): 1231-1240.
- Iino, T., H. Tamaki, *et al.* (2013). "*Candidatus* Methanogranum caenicola: a novel methanogen from the anaerobic digested sludge, and proposal of *Methanomassiliicoccaceae* fam. nov. and *Methanomassiliicoccales* ord. nov., for a methanogenic lineage of the class *Thermoplasmata*." *Microbes Environ* 28(2): 244-250.
- Ikemura, T. and M. Nomura (1977). "Expression of spacer tRNA genes in ribosomal RNA transcription units carried by hybrid Col E1 plasmids in *E. coli*." *Cell* 11(4): 779-793.
- Ilag, L. L., D. Jahn, *et al.* (1991). "The *Escherichia coli hemL* gene encodes glutamate 1-semialdehyde aminotransferase." *J Bacteriol* 173(11): 3408-3413.
- Imachi, H., S. Sakai, *et al.* (2009). "*Methanofollis ethanolicus* sp. nov., an ethanol-utilizing methanogen isolated from a lotus field." *Int J Syst Evol Microbiol* 59(Pt 4): 800-805.

- Immig, I., D. Demeyer, *et al.* (1996). "Attempts to induce reductive acetogenesis into a sheep rumen." *Arch Tierernähr* 49(4): 363-370.
- Irbis, C. and K. Ushida (2004). "Detection of methanogens and proteobacteria from a single cell of rumen ciliate protozoa." *J Gen Appl Microbiol* 50(4): 203-212.
- Iverson, V., R. M. Morris, *et al.* (2012). "Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota." *Science* 335(6068): 587-590.
- Jansen, K., E. Stupperich, *et al.* (1982). "Carbohydrate synthesis from acetyl CoA in the autotroph *Methanobacterium thermoautotrophicum*." *Arch Microbiol* 132(4): 355-364.
- Janssen, P. H. and M. Kirs (2008). "Structure of the archaeal community of the rumen." *Appl Environ Microbiol* 74(12): 3619-3625.
- Janssen, P. H. (2010). "Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics." *Anim Feed Sci Technol* 160(1-2): 1-22.
- Janssen, P. H. and M. Kirs (2008). "Structure of the archaeal community of the rumen." *Appl Environ Microbiol* 74(12): 3619-3625.
- Jarrell, K. F., G. M. Jones, *et al.* (2010). "S-layer glycoproteins and flagellins: reporters of archaeal posttranslational modifications." *Archaea* 2010.
- Jarvis, G. N., C. Strömpl, *et al.* (2000). "Isolation and identification of ruminal methanogens from grazing cattle." *Curr Microbiol* 40(5): 327-332.
- Jayanegara, A., E. Wina, *et al.* (2014). "Meta-analysis on methane mitigating properties of saponin-rich sources in the rumen: Influence of addition levels and plant sources." *Asian-Australas J Anim Sci* 27(10): 1426-1435.
- Jenkins, J., V. E. Shevchik, *et al.* (2004). "The crystal structure of pectate lyase Pel9A from *Erwinia chrysanthemi*." *J Biol Chem* 279(10): 9139-9145.
- Jeyanathan, J. (2010). Investigation of rumen methanogens in New Zealand livestock. Philosophy in Animal Science Thesis, Massey University.
- Jeyanathan, J., M. Kirs, *et al.* (2011). "Methanogen community structure in the rumens of farmed sheep, cattle and red deer fed different diets." *FEMS Microbiol Ecol* 76(2): 311-326.

- Jin, W., Y. F. Cheng, *et al.* (2014). "Discovery of a novel rumen methanogen in the anaerobic fungal culture and its distribution in the rumen as revealed by real-time PCR." *BMC Microbiol* 14(1): 104.
- Jindou, S., J. M. Brulc, *et al.* (2008). "Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium *Ruminococcus flavefaciens*." *FEMS Microbiol Lett* 285(2): 188-194.
- Joblin, K. N. (1999). "Ruminal acetogens and their potential to lower ruminant methane emissions." *Aust J Agric Res* 50(8): 1307-1314.
- Jofre, E., A. Lagares, *et al.* (2004). "Disruption of dTDP-rhamnose biosynthesis modifies lipopolysaccharide core, exopolysaccharide production, and root colonization in *Azospirillum brasilense*." *FEMS Microbiol Lett* 231(2): 267-275.
- Johansson, L., G. Gafvelin, *et al.* (2005). "Selenocysteine in proteins-properties and biotechnological use." *Biochim Biophys Acta* 1726(1): 1-13.
- Johnson, K. A. and D. E. Johnson (1995). "Methane emissions from cattle." *J Anim Sci* 73(8): 2483-2492.
- Johnson, P. F. and J. Abelson (1983). "The yeast tRNA<sup>Tyr</sup> gene intron is essential for correct modification of its tRNA product." *Nature* 302: 681 - 687.
- Jonah, M. and J. A. Erwin (1971). "The lipids of membraneous cell organelles isolated from the ciliate, *Tetrahymena pyriformis*." *Biochim Biophys Acta* 231(1): 80-92.
- Jones, D. T., W. R. Taylor, *et al.* (1992). "The rapid generation of mutation data matrices from protein sequences." *Comput Appl Biosci* 8(3): 275-282.
- Jongsareejit, B., R. N. Rahman, *et al.* (1997). "Gene cloning, sequencing and enzymatic properties of glutamate synthase from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1." *Mol Gen Genet* 254(6): 635-642.
- Jordan, P. M., S. D. Thomas, *et al.* (1988). "Purification, crystallization and properties of porphobilinogen deaminase from a recombinant strain of *Escherichia coli* K12." *Biochem. J* 254: 427-435.

- Kaback, H. R., M. Sahin-Toth, *et al.* (2001). "The kamikaze approach to membrane transport." *Nat Rev Mol Cell Biol* 2(8): 610-620.
- Kaesler, B. and P. Schonheit (1989). "The role of sodium ions in methanogenesis. Formaldehyde oxidation to CO<sub>2</sub> and 2H<sub>2</sub> in methanogenic bacteria is coupled with primary electrogenic Na<sup>+</sup> translocation at a stoichiometry of 2-3 Na<sup>+</sup>/CO<sub>2</sub>." *Eur J Biochem* 184(1): 223-232.
- Kameda, Y., S. Oira, *et al.* (1974). "Antitumor activity of bacillus natto. V. Isolation and characterization of surfactin in the culture medium of *Bacillus natto* KMD 2311." *Chem Pharm Bull (Tokyo)* 22(4): 938-944.
- Kaminski, L. and J. Eichler (2014). "Haloferax volcanii *N*-glycosylation: delineating the pathway of dTDP-rhamnose biosynthesis." *PloS one* 9(5): e97441.
- Kandler, O. and H. König (1978). "Chemical composition of the peptidoglycan-free cell walls of methanogenic bacteria." *Arch Microbiol* 118(2): 141-152.
- Kandler, O. and H. König (1998). "Cell wall polymers in Archaea (Archaeobacteria)." *Cell Mol Life Sci* 54(4): 305-308.
- Kanehisa, M. and S. Goto (2000). "KEGG: kyoto encyclopedia of genes and genomes." *Nucleic Acids Res* 28(1): 27-30.
- Karl, T. R. and K. E. Trenberth (2003). "Modern global climate change." *Science* 302(5651): 1719-1723.
- Kaster, A. K., J. Moll, *et al.* (2011). "Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea." *Proc Natl Acad Sci U S A* 108(7): 2981-2986.
- Kato, N., H. Yurimoto, *et al.* (2006). "The physiological role of the ribulose monophosphate pathway in bacteria and archaea." *Biosci Biotechnol Biochem* 70(1): 10-21.
- Kawamukai, M., R. Utsumi, *et al.* (1991). "Nucleotide sequence and characterization of the *sfs1* gene: *sfs1* is involved in CRP\*-dependent *mal* gene expression in *Escherichia coli*." *J Bacteriol* 173(8): 2644-2648.

- Kearse, M., R. Moir, *et al.* (2012). "Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data." *Bioinformatics* 28(12): 1647-1649.
- Keller, M., G. Burkard, *et al.* (1980). "Transfer RNA genes associated with the 16S and 23S rRNA genes of *Euglena* chloroplast DNA." *Biochem Biophys Res Commun* 95(1): 47-54.
- Kelly, W. J., S. C. Leahy, *et al.* (2014). "The complete genome sequence of the rumen methanogen *Methanobacterium formicicum* BRM9." *Stand Genomic Sci* 9: 15.
- Kelly, W. J., D. M. Pacheco, *et al.* (2016). "The complete genome sequence of the rumen methanogen *Methanobrevibacter millerae* SM9." *Stand Genomic Sci* unpublished(unknown issue): unknown page.
- Kelly, W. J., D. Li, *et al.* (2016). "Draft Genome Sequence of the Rumen Methanogen *Methanobrevibacter olleyae* YLM1." *Genome Announc* 4(2).
- Keltjens, J. T. and G. D. Vogels (1993). Conversion of methanol and methylamines to methane and carbon dioxide. *Methanogenesis*, Springer: 253-303.
- Kenklies, J., R. Ziehn, *et al.* (1999). "Proline biosynthesis from L-ornithine in *Clostridium sticklandii*: purification of  $\Delta 1$ -pyrroline-5-carboxylate reductase, and sequence and expression of the encoding gene, *proC*." *Microbiology* 145 ( Pt 4): 819-826.
- Khakh, B. S. and G. Burnstock (2009). "The double life of ATP." *Sci Am* 301(6): 84-90, 92.
- Khan, H., S. H. Flint, *et al.* (2013). "Determination of the mode of action of enterolysin A, produced by *Enterococcus faecalis* B9510." *J Appl Microbiol* 115(2): 484-494.
- Khelaifia, S., M. Garibal, *et al.* (2014). "Draft genome sequence of a human-associated isolate of *Methanobrevibacter arboriphilicus*, the lowest-G+C-content archaeon." *Genome Announc* 2(1).
- Khelaifia, S., M. Garibal, *et al.* (2014). "Draft genome sequencing of *Methanobrevibacter oralis* strain JMR01, isolated from the human intestinal microbiota." *Genome Announc* 2(1).
- Khomyakova, M., Ö. Bükmez, *et al.* (2011). "A methylaspartate cycle in haloarchaea." *Science* 331(6015): 334-337.

Kiehl, J. T. and K. E. Trenberth (1997). "Earth's annual global mean energy budget." *Bull Am Meteorol Soc* 78(2): 197-208.

Kim, C. (2012). Identification of rumen methanogens, characterization of substrate requirements and measurement of hydrogen thresholds. Master in Microbiology, Massey University.

Kim, S. Y., K. S. Ju, *et al.* (2012). "Different biosynthetic pathways to fosfomycin in *Pseudomonas syringae* and *Streptomyces* species." *Antimicrob Agents Chemother* 56(8): 4175-4183.

Kimura, A., K. T. Mountzouros, *et al.* (1990). "*Bordetella pertussis* filamentous hemagglutinin: evaluation as a protective antigen and colonization factor in a mouse respiratory infection model." *Infect Immun* 58(1): 7-16.

Kimura, M. (1980). "A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences." *J Mol Evol* 16(2): 111-120.

Kirby, J., J. C. Martin, *et al.* (1997). "Dockerin-like sequences in cellulases and xylanases from the rumen cellulolytic bacterium *Ruminococcus flavefaciens*." *FEMS Microbiol Lett* 149(2): 213-219.

Kitts, C. L., D. P. Cunningham, *et al.* (1994). "Isolation of three hexahydro-1,3,5-trinitro-1,3,5-triazine-degrading species of the family Enterobacteriaceae from nitramine explosive-contaminated soil." *Appl Environ Microbiol* 60(12): 4608-4611.

Kjems, J. and R. A. Garrett (1987). "Novel expression of the ribosomal RNA genes in the extreme thermophile and archaeobacterium *Desulfurococcus mobilis*." *EMBO J* 6(11): 3521.

Kloda, A. and B. Martinac (2001). "Mechanosensitive channel of *Thermoplasma*, the cell wall-less archaea: cloning and molecular characterization." *Cell Biochem Biophys* 34(3): 321-347.

Kneidinger, B., M. Graninger, *et al.* (2001). "Biosynthesis of nucleotide-activated D-glycero-D-manno-heptose." *J Biol Chem* 276(24): 20935-20944.

Kolberg, M., K. R. Strand, *et al.* (2004). "Structure, function, and mechanism of ribonucleotide reductases." *BBA-Protein Proteomics* 1699(1-2): 1-34.

- König, H. and O. Kandler (1979). "The amino acid sequence of the peptide moiety of the pseudomurein from *Methanobacterium thermoautotrophicum*." Arch Microbiol 121(3): 271-275.
- Koonin, E. V., K. S. Makarova, *et al.* (2001). "Horizontal gene transfer in prokaryotes: quantification and classification." Annu Rev Microbiol 55: 709-742.
- Koval, S. F. and K. F. Jarrell (1987). "Ultrastructure and biochemistry of the cell wall of *Methanococcus voltae*." J Bacteriol 169(3): 1298-1306.
- Krogh, A., B. Larsson, *et al.* (2001). "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes." J Mol Biol 305(3): 567-580.
- Kruskal, W. H. and W. A. Wallis (1952). "Use of ranks in one-criterion variance analysis." J Amer Statist. Assoc 47(260): 583-621.
- Krzycki, J. A. (2004). "Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases." Curr Opin Chem Biol 8(5): 484-491.
- Kühn, W., K. Fiebig, *et al.* (1983). "Distribution of cytochromes in methanogenic bacteria." FEMS Microbiol Lett 20(3): 407-410.
- Kubo, I., K. Nihei, *et al.* (2003). "Antibacterial action of anacardic acids against methicillin resistant *Staphylococcus aureus* (MRSA)." J Agric Food Chem 51(26): 7624-7628.
- Kulkarni, G., D. M. Kridelbaugh, *et al.* (2009). "Hydrogen is a preferred intermediate in the energy-conserving electron transport chain of *Methanosarcina barkeri*." Proc Natl Acad Sci U S A 106(37): 15915-15920.
- Kullen, M. J. and T. R. Klaenhammer (1999). "Identification of the pH-inducible, proton-translocating F<sub>1</sub>F<sub>0</sub>-ATPase (*atpBEFHAGDC*) operon of *Lactobacillus acidophilus* by differential display: gene structure, cloning and characterization." Mol Microbiol 33(6): 1152-1161.
- Kunkel, A., J. A. Vorholt, *et al.* (1998). "An *Escherichia coli* hydrogenase-3-type hydrogenase in methanogenic archaea." Eur J Biochem 252(3): 467-476.
- Kuzuyama, T. (2002). "Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units." Biosci Biotechnol Biochem 66(8): 1619-1627.

- L'Haridon, S., M. Chalopin, *et al.* (2014). "*Methanococcoides vulcani* sp. nov., a marine methylotrophic methanogen that uses betaine, choline and *N,N*-dimethylethanolamine for methanogenesis, isolated from a mud volcano, and emended description of the genus *Methanococcoides*." *Int J Syst Evol Microbiol* 64(Pt 6): 1978-1983.
- Lambie, S. C., W. J. Kelly, *et al.* (2015). "The complete genome sequence of the rumen methanogen *Methanosarcina barkeri* CM1." *Stand Genomic Sci* 10: 57.
- Lamed, R., J. Naimark, *et al.* (1987). "Specialized cell surface structures in cellulolytic bacteria." *J Bacteriol* 169(8): 3792-3800.
- Lang, K., J. Schuldes, *et al.* (2015). "New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genome analysis of "*Candidatus* Methanoplasma termitum"." *Appl Environ Microbiol* 81(4): 1338-1352.
- Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." *Nat Methods* 9(4): 357-359.
- Lanigan, G. W., A. L. Payne, *et al.* (1978). "Antimethanogenic drugs and *Heliotropium europaeum* poisoning in penned sheep." *Crop Pasture Sci* 29(6): 1281-1292.
- Larimer, F. W. (1987). "Cleavage by *ApaI* is inhibited by overlapping dcm methylation." *Nucleic Acids Res* 15(21): 9087.
- Larsen, T. M., S. K. Boehlein, *et al.* (1999). "Three-dimensional structure of *Escherichia coli* asparagine synthetase B: a short journey from substrate to product." *Biochemistry* 38(49): 16146-16157.
- Lazazzera, B. A. (2000). "Quorum sensing and starvation: signals for entry into stationary phase." *Curr Opin Microbiol* 3(2): 177-182.
- Leadbetter, J. R. and J. A. Breznak (1996). "Physiological ecology of *Methanobrevibacter cuticularis* sp. nov. and *Methanobrevibacter curvatus* sp. nov., isolated from the hindgut of the termite *Reticulitermes flavipes*." *Appl Environ Microbiol* 62(10): 3620-3631.
- Leadbetter, J. R., L. D. Crosby, *et al.* (1998). "*Methanobrevibacter filiformis* sp. nov., A filamentous methanogen from termite hindguts." *Arch Microbiol* 169(4): 287-292.



- Lê, S., J. Josse, *et al.* (2008). "FactoMineR: an R package for multivariate analysis." *J Stat Softw* 25(1): 1-18.
- Leahy, S. C., W. J. Kelly, *et al.* (2010). "The genome sequence of the rumen methanogen *Methanobrevibacter ruminantium* reveals new possibilities for controlling ruminant methane emissions." *PloS one* 5(1): e8926.
- Leahy, S. C., W. J. Kelly, *et al.* (2013). "Genome sequencing of rumen bacteria and archaea and its application to methane mitigation strategies." *Animal* 7 Suppl 2: 235-243.
- Leahy, S. C., W. J. Kelly, *et al.* (2013). "The complete genome sequence of *Methanobrevibacter* sp. AbM4." *Stand Genomic Sci* 8(2): 215-227.
- Lee, J. H., S. Kumar, *et al.* (2013). "*Methanobrevibacter boviskoreani* sp. nov., isolated from the rumen of Korean native cattle." *Int J Syst Evol Microbiol* 63(Pt 11): 4196-4201.
- Lee, J. H., M. S. Rhee, *et al.* (2013). "Genome sequence of *Methanobrevibacter* sp. strain JH1, isolated from rumen of Korean native cattle." *Genome Announc* 1(1).
- Lee, J. K. and K. N. Houk (1997). "A proficient enzyme revisited: the predicted mechanism for orotidine monophosphate decarboxylase." *Science* 276(5314): 942-945.
- Legendre, P. and E. D. Gallagher (2001). "Ecologically meaningful transformations for ordination of species data." *Oecol* 129(2): 271-280.
- Legendre, P., J. Oksanen, *et al.* (2011). "Testing the significance of canonical axes in redundancy analysis." *Methods Ecol Evol* 2(3): 269-277.
- Leigh, J. A. (2000). "Nitrogen fixation in methanogens: the archaeal perspective." *Curr Issues Mol Biol* 2(4): 125-131.
- Levis, R., P. Dunsmuir, *et al.* (1980). "Terminal repeats of the *Drosophila* transposable element copia: nucleotide sequence and genomic organization." *Cell* 21(2): 581-588.
- Li, X., Y. Yang, *et al.* (2014). "Comparative proteomics analyses of *Kobresia pygmaea* adaptation to environment along an elevational gradient on the central Tibetan Plateau." *PloS one* 9(6): e98410.

- Lie, T. J., K. C. Costa, *et al.* (2012). "Essential anaplerotic role for the energy-converting hydrogenase Eha in hydrogenotrophic methanogenesis." *Proc Natl Acad Sci U S A* 109(38): 15473-15478.
- Lieb, M. (1991). "Spontaneous mutation at a 5-methylcytosine hotspot is prevented by very short patch (VSP) mismatch repair." *Genetics* 128(1): 23-27.
- Lienard, T., B. Becher, *et al.* (1996). "Sodium ion translocation by N<sup>5</sup>-methyltetrahydromethanopterin: coenzyme M methyltransferase from *Methanosarcina mazei* Gö1 reconstituted in ether lipid liposomes." *Eur J Biochem* 239(3): 857-864.
- Lin, B., J. H. Wang, *et al.* (2013). "*In vitro* rumen fermentation and methane production are influenced by active components of essential oils combined with fumarate." *J Anim Physiol Anim Nutr (Berl)* 97(1): 1-9.
- Lin, C., L. Raskin, *et al.* (1997). "Microbial community structure in gastrointestinal tracts of domestic animals: Comparative analyses using rRNA-targeted oligonucleotide probes." *FEMS Microbiol Ecol* 22(4): 281-294.
- Liu, Y. and W. B. Whitman (2008). "Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea." *Ann N Y Acad Sci* 1125: 171-189.
- Liu, Y., R. H. White, *et al.* (2010). "*Methanococci* use the diaminopimelate aminotransferase (DapL) pathway for lysine biosynthesis." *J Bacteriol* 192(13): 3304-3310.
- Lobo, S. A., A. Brindley, *et al.* (2009). "Functional characterization of the early steps of tetrapyrrole biosynthesis and modification in *Desulfovibrio vulgaris* Hildenborough." *Biochem J* 420(2): 317-325.
- Loenen, W. A. M. (2006). "S-Adenosylmethionine: jack of all trades and master of everything?" *Biochem Soc Trans* 34.
- Longstaff, D. G., S. K. Blight, *et al.* (2007). "*In vivo* contextual requirements for UAG translation as pyrrolysine." *Mol Microbiol* 63(1): 229-241.
- Lowe, T. M. and S. R. Eddy (1997). "tR NAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence." *Nucleic Acids Res* 25(5): 955-964.

- Lu, G. and E. N. Moriyama (2004). "Vector NTI, a balanced all-in-one sequence analysis suite." *Brief Bioinform* 5(4): 378-388.
- Lund, E., J. E. Dahlberg, *et al.* (1976). "Transfer RNA genes between 16S and 23S rRNA genes in rRNA transcription units of *E. coli*." *Cell* 7(2): 165-177.
- Luo, D., S. Ganesh, *et al.* (2014). "Package 'predictmeans'."
- Luton, P. E., J. M. Wayne, *et al.* (2002). "The *mcrA* gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill." *Microbiology* 148(Pt 11): 3521-3530.
- Mackie, R. I. and S. Heath (1979). "Enumeration and isolation of lactate-utilizing bacteria from the rumen of sheep." *Appl Environ Microbiol* 38(3): 416-421.
- MacPherson, L. (2014). "Agricultural production statistics: June 2014 (final)." Retrieved 28/07, 2015, from [http://www.stats.govt.nz/browse\\_for\\_stats/industry\\_sectors/agriculture-horticulture-forestry/AgriculturalProduction\\_final\\_HOTPJun14final.aspx](http://www.stats.govt.nz/browse_for_stats/industry_sectors/agriculture-horticulture-forestry/AgriculturalProduction_final_HOTPJun14final.aspx).
- Maeder, D. L., I. Anderson, *et al.* (2006). "The *Methanosarcina barkeri* genome: comparative analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* reveals extensive rearrangement within methanosarcinal genomes." *J Bacteriol* 188(22): 7922-7931.
- Magni, G., G. Orsomando, *et al.* (2006). "Structural and functional properties of NAD kinase, a key enzyme in NADP biosynthesis." *Mini Rev Med Chem* 6(7): 739-746.
- Magoc, T., D. Wood, *et al.* (2013). "EDGE-pro: estimated degree of gene expression in prokaryotic genomes." *Evol Bioinform Online* 9: 127-136.
- Mahapatra, A., A. Patel, *et al.* (2006). "Characterization of a *Methanosarcina acetivorans* mutant unable to translate UAG as pyrrolysine." *Mol Microbiol* 59(1): 56-66.
- Mahillon, J. and M. Chandler (1998). "Insertion sequences." *Microbiol Mol Biol Rev* 62(3): 725-774.
- Maizel, J. V., A. A. Benson, *et al.* (1956). "Identification of phosphoryl choline as an important constituent of plant sap." *Plant Physiol* 31(5): 407-408.
- Major, T. A., Y. Liu, *et al.* (2010). "Characterization of energy-conserving hydrogenase B in *Methanococcus maripaludis*." *J Bacteriol* 192(15): 4022-4030.

- Makarova, K. S., L. Aravind, *et al.* (2011). "Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems." *Biol Direct* 6: 38.
- Makarova, K. S., D. H. Haft, *et al.* (2011). "Evolution and classification of the CRISPR-Cas systems." *Nat Rev Microbiol* 9(6): 467-477.
- Makkar, H. P., M. Blümmel, *et al.* (1995). "In vitro effects of and interactions between tannins and saponins and fate of tannins in the rumen." *J Sci Food Agric* 69(4): 481-493.
- Mander, G. J., A. J. Pierik, *et al.* (2004). "Two distinct heterodisulfide reductase-like enzymes in the sulfate-reducing archaeon *Archaeoglobus profundus*." *Eur J Biochem* 271(6): 1106-1116.
- Margolies, M. N. and R. F. Goldberger (1966). "Isolation of the fourth enzyme (isomerase) of histidine biosynthesis from *Salmonella typhimurium*." *J Biol Chem* 241(14): 3262-3269.
- Margulies, M., M. Egholm, *et al.* (2005). "Genome sequencing in microfabricated high-density picolitre reactors." *Nature* 437(7057): 376-380.
- Markham, G. D., E. W. Hafner, *et al.* (1980). "S-Adenosylmethionine synthetase from *Escherichia coli*." *J Biol Chem* 255(19): 9082-9092.
- Marolewski, A., J. M. Smith, *et al.* (1994). "Cloning and characterization of a new purine biosynthetic enzyme: a non-folate glycinamide ribonucleotide transformylase from *E. coli*." *Biochemistry* 33(9): 2531-2537.
- Marraffini, L. A. and E. J. Sontheimer (2008). "CRISPR interference limits horizontal gene transfer in *Staphylococci* by targeting DNA." *Science* 322(5909): 1843-1845.
- Marraffini, L. A. and E. J. Sontheimer (2010). "CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea." *Nat Rev Genet* 11(3): 181-190.
- Martin, S. A. (1998). "Manipulation of ruminal fermentation with organic acids: a review." *J Anim Sci* 76(12): 3123-3132.
- Mashhadi, Z., H. Xu, *et al.* (2010). "Archaeal RibL: a new FAD synthetase that is air sensitive." *Biochemistry* 49(40): 8748-8755.
- Mashhadi, Z., H. Zhang, *et al.* (2008). "Identification and characterization of an archaeon-specific riboflavin kinase." *J Bacteriol* 190(7): 2615-2618.

- Matarazzo, F., A. C. Ribeiro, *et al.* (2012). "The domain Archaea in human mucosal surfaces." *Clin Microbiol Infect* 18(9): 834-840.
- May, M. S. and S. Hattman (1975). "Analysis of bacteriophage deoxyribonucleic acid sequences methylated by host-and R-factor-controlled enzymes." *J Bacteriol* 123(2): 768-770.
- McAllister, T. A. and C. J. Newbold (2008). "Redirecting rumen fermentation to reduce methanogenesis." *Aust J Exp Agric* 48(2): 7-13.
- McArthur, J. M. and J. E. Miltimore (1961). "Rumen gas analysis by gas-solid chromatography." *Can J Anim Sci* 41(2): 187-196.
- McClure, R., D. Balasubramanian, *et al.* (2013). "Computational analysis of bacterial RNA-Seq data." *Nucleic Acids Res* 41(14): e140.
- McGinn, S. M., K. A. Beauchemin, *et al.* (2004). "Methane emissions from beef cattle: Effects of monensin, sunflower oil, enzymes, yeast, and fumaric acid." *J Anim Sci* 82(11): 3346-3356.
- McDonald, M. J., C. H. Chou, *et al.* (2015). "The evolutionary dynamics of tRNA-gene copy number and codon-use in *E. coli*." *BMC Evol Biol* 15: 163.
- McIntosh, E. M. and R. H. Haynes (1986). "Sequence and expression of the dCMP deaminase gene (DCD1) of *Saccharomyces cerevisiae*." *Mol Cell Biol* 6(5): 1711-1721.
- McMillan, D. G. G., S. A. Ferguson, *et al.* (2011). "A<sub>1</sub>A<sub>0</sub>-ATP synthase of *Methanobrevibacter ruminantium* couples sodium ions for ATP synthesis under physiological conditions." *J Biol Chem* 286(46): 39882-39892.
- Mendez-Garcia, C., V. Mesa, *et al.* (2014). "Microbial stratification in low pH oxic and suboxic macroscopic growths along an acid mine drainage." *ISME J* 8(6): 1259-1274.
- Mengin-Lecreulx, D. and J. van Heijenoort (1996). "Characterization of the essential gene *glmM* encoding phosphoglucosamine mutase in *Escherichia coli*." *J Biol Chem* 271(1): 32-39.
- Messenger, L. J. and H. Zalkin (1979). "Glutamine phosphoribosylpyrophosphate amidotransferase from *Escherichia coli*. Purification and properties." *J Biol Chem* 254(9): 3382-3392.

- Meyer, E., N. J. Leonard, *et al.* (1992). "Purification and characterization of the *purE*, *purK*, and *purC* gene products: identification of a previously unrecognized energy requirement in the purine biosynthetic pathway." *Biochemistry* 31(21): 5022-5032.
- Middleton, B. (1974). "The kinetic mechanism and properties of the cytoplasmic acetoacetyl-coenzyme A thiolase from rat liver." *Biochem J* 139(1): 109-121.
- Millar, G., A. Lewendon, *et al.* (1986). "The cloning and expression of the *aroL* gene from *Escherichia coli* K12. Purification and complete amino acid sequence of shikimate kinase II, the *aroL*-gene product." *Biochem. J* 237: 427-437.
- Miller, M. B., K. Skorupski, *et al.* (2002). "Parallel quorum sensing systems converge to regulate virulence in *Vibrio cholerae*." *Cell* 110(3): 303-314.
- Miller, T. L. (2001). *Methanobrevibacter*. *Bergey's Manual of Systematics of Archaea and Bacteria*. W. B. Whitman. Online, Wiley.
- Miller, T. L. and C. Lin (2002). "Description of *Methanobrevibacter gottschalkii* sp. nov., *Methanobrevibacter thaueri* sp. nov., *Methanobrevibacter woesei* sp. nov. and *Methanobrevibacter wolinii* sp. nov." *Int J Syst Evol Microbiol* 52(Pt 3): 819-822.
- Miller, T. L., M. J. Wolin, *et al.* (1982). "Isolation of *Methanobrevibacter smithii* from human feces." *Appl Environ Microbiol* 43(1): 227-232.
- Ministry of the Environment. (2015). New Zealand's Greenhouse Gas Inventory 1990-2013. Ministry of the Environment. Wellington, New Zealand.
- Mitchell, A. D., A. Chappell, *et al.* (1979). "Metabolism of betaine in the ruminant." *J Anim Sci* 49(3): 764-774.
- Moe, P. W. and H. F. Tyrrell (1979). "Methane production in dairy cows." *J Dairy Sci* 62(10): 1583-1586.
- Moore, D. P. and J. F. Thompson (1967). "Methionine biosynthesis from *S*-methylcysteine by thiomethyl transfer in *Neurospora*." *Biochem Biophys Res Commun* 28(3): 474-479.
- Morgan, E. A., T. Ikemura, *et al.* (1977). "Identification of spacer tRNA genes in individual ribosomal RNA transcription units of *Escherichia coli*." *Proc Natl Acad Sci U S A* 74(7): 2710-2714.

- Morgavi, D. P., W. J. Kelly, *et al.* (2012). "Rumen microbial (meta)genomics and its application to ruminant production." *Animal* 7(s1): 184-201
- Morgavi, D. P., C. Martin, *et al.* (2012). "Rumen protozoa and methanogenesis: Not a simple cause-effect relationship." *Br J Nutr* 107(3): 388-397.
- Morii, H. and Y. Koga (2003). "CDP-2,3-Di-*O*-geranylgeranyl-*sn*-glycerol:L-serine *O*-archaetidyltransferase (archaetidylserine synthase) in the methanogenic archaeon *Methanothermobacter thermautotrophicus*." *J Bacteriol* 185(4): 1181-1189.
- Morii, H., M. Nishihara, *et al.* (2000). "CTP:2,3-di-*O*-geranylgeranyl-*sn*-glycero-1-phosphate cytidyltransferase in the methanogenic archaeon *Methanothermobacter thermoautotrophicus*." *J Biol Chem* 275(47): 36568-36574.
- Morrison, S. D., S. A. Roberts, *et al.* (2008). "A new use for a familiar fold: the X-ray crystal structure of GTP-bound GTP cyclohydrolase III from *Methanocaldococcus jannaschii* reveals a two metal ion catalytic mechanism." *Biochemistry* 47(1): 230-242.
- Morton, T. A., J. A. Runquist, *et al.* (1991). "The primary structure of the subunits of carbon monoxide dehydrogenase/acetyl-CoA synthase from *Clostridium thermoaceticum*." *J Biol Chem* 266(35): 23824-23828.
- Morvan, B., F. Bonnemoy, *et al.* (1996). "Quantitative determination of H<sub>2</sub>-utilizing acetogenic and sulfate-reducing bacteria and methanogenic archaea from digestive tract of different mammals." *Curr Microbiol* 32(3): 129-133.
- Moss, A. R., J. P. Jouany, *et al.* (2000). "Methane production by ruminants: its contribution to global warming." *Ann Zootech* 49: 231-253.
- Mount, D. W. (2008). "Using BLOSUM in sequence alignments." *CSH Protoc* 2008: pdb top39.
- Muller, E., K. Fahlbusch, *et al.* (1981). "Formation of *N,N*-dimethylglycine, acetic acid, and butyric acid from betaine by *Eubacterium limosum*." *Appl Environ Microbiol* 42(3): 439-445.
- Muller, V., M. Blaut, *et al.* (1986). "Utilization of methanol plus hydrogen by *Methanosarcina barkeri* for methanogenesis and growth." *Appl Environ Microbiol* 52(2): 269-274.

- Muller, V., C. Ruppert, *et al.* (1999). "Structure and function of the A<sub>1</sub>A<sub>0</sub>-ATPases from methanogenic Archaea." *J Bioenerg Biomembr* 31(1): 15-27.
- Muller, V. (2003). "Energy conservation in acetogenic bacteria." *Appl Environ Microbiol* 69(11): 6345-6353.
- Musfeldt, M., M. Selig, *et al.* (1999). "Acetyl coenzyme A synthetase (ADP forming) from the hyperthermophilic archaeon *Pyrococcus furiosus*: identification, cloning, separate expression of the encoding genes, *acdAI* and *acdBI*, in *Escherichia coli*, and in vitro reconstitution of the active heterotetrameric enzyme from its recombinant subunits." *J Bacteriol* 181(18): 5885-5888.
- Myllykallio, H., G. Lipowski, *et al.* (2002). "An alternative flavin-dependent mechanism for thymidylate synthesis." *Science* 297(5578): 105-107.
- Nagata, Y., K. Tanaka, *et al.* (1999). "Occurrence of D-amino acids in a few archaea and dehydrogenase activities in hyperthermophile *Pyrobaculum islandicum*." *Biochim Biophys Acta* 1435(1-2): 160-166.
- Narindrasorasak, S. and W. A. Bridger (1977). "Phosphoenolpyruvate synthetase of *Escherichia coli*: molecular weight, subunit composition, and identification of phosphohistidine in phosphoenzyme intermediate." *J Biol Chem* 252(10): 3121-3127.
- Naumann, E., H. Hippe, *et al.* (1983). "Betaine: New Oxidant in the Stickland reaction and methanogenesis from betaine and L-Alanine by a *Clostridium sporogenes*-*Methanosarcina barkeri* Coculture." *Appl Environ Microbiol* 45(2): 474-483.
- Nawrocki, E. P., S. W. Burge, *et al.* (2015). "Rfam 12.0: updates to the RNA families database." *Nucleic Acids Res* 43(Database issue): D130-137.
- Nawrocki, E. P. and S. R. Eddy (2013). "Infernal 1.1: 100-fold faster RNA homology searches." *Bioinformatics* 29(22): 2933-2935.
- Neill, A. R., D. W. Grime, *et al.* (1978). "Conversion of choline methyl groups through trimethylamine into methane in the rumen." *Biochem J* 170(3).
- Nejat, N., G. Vadamalai, *et al.* (2013). "'*Candidatus Phytoplasma malaysianum*', a novel taxon associated with virescence and phyllody of Madagascar periwinkle (*Catharanthus roseus*)." *Int J Syst Evol Microbiol* 63(Pt 2): 540-548.



- Neuwald, A. F. and G. V. Stauffer (1985). "DNA sequence and characterization of the *Escherichia coli serB* gene." *Nucleic Acids Res* 13(19): 7025-7039.
- Newbold, C. J., S. Lopez, *et al.* (2005). "Propionate precursors and other metabolic intermediates as possible alternative electron acceptors to methanogenesis in ruminal fermentation in vitro." *Br J Nutr* 94(1): 27-35.
- Ng, F., S. Kittelmann, *et al.* (2016). "An adhesin from hydrogen-utilizing rumen methanogen *Methanobrevibacter ruminantium* M1 binds a broad range of hydrogen-producing microorganisms." *Environ Microbiol.*
- Nielsen, F. S., P. S. Andersen, *et al.* (1996). "The B form of dihydroorotate dehydrogenase from *Lactococcus lactis* consists of two different subunits, encoded by the *pyrDb* and *pyrK* genes, and contains FMN, FAD, and [FeS] redox centers." *J Biol Chem* 271(46): 29359-29365.
- Nishihara, M. and Y. Koga (1997). "Purification and properties of *sn*-glycerol-1-phosphate dehydrogenase from *Methanobacterium thermoautotrophicum*: characterization of the biosynthetic enzyme for the enantiomeric glycerophosphate backbone of ether polar lipids of Archaea." *J Biochem* 122(3): 572-576.
- Nishimura, Y. and T. Eguchi (2006). "Biosynthesis of archaeal membrane lipids: digeranylgeranylglycerophospholipid reductase of the thermoacidophilic archaeon *Thermoplasma acidophilum*." *J Biochem* 139(6): 1073-1081.
- Niu, Y. H. (2014). Effect of different molecular weight of tannins from alpine plants on in vitro methane production in yaks [Thesis Abstract]. Master.
- Noel, S. J., O. Hojberg, *et al.* (2016). "Draft genome sequence of "*Candidatus* Methanomethylophilus" sp. 1R26, enriched from bovine rumen, a methanogenic archaeon belonging to the *Methanomassiliicoccales* Order." *Genome Announc* 4(1).
- Norager, S., K. F. Jensen, *et al.* (2002). "*E. coli* dihydroorotate dehydrogenase reveals structural and functional distinctions between different classes of dihydroorotate dehydrogenases." *Structure* 10(9): 1211-1223.
- Noll, K. M., K. L. Rinehart, *et al.* (1986). "Structure of component B (7-mercaptoheptanoylthreonine phosphate) of the methylcoenzyme M methylreductase system of *Methanobacterium thermoautotrophicum*." *PNAS* 83(12): 4238-4242.

- Nusslein, B., K. J. Chin, *et al.* (2001). "Evidence for anaerobic syntrophic acetate oxidation during methane production in the profundal sediment of subtropical Lake Kinneret (Israel)." *Environ Microbiol* 3(7): 460-470.
- O'Sullivan, O., M. Begley, *et al.* (2011). "Further identification of novel lantibiotic operons using LanM-based genome mining." *Probiotics Antimicrob Proteins* 3(1): 27-40.
- Ochman, H., J. G. Lawrence, *et al.* (2000). "Lateral gene transfer and the nature of bacterial innovation." *Nature* 405(6784): 299-304.
- Ogawa, H., T. Gomi, *et al.* (1987). "Amino acid sequence of S-adenosyl-L-homocysteine hydrolase from rat liver as derived from the cDNA sequence." *PNAS* 84(3): 719-723.
- Ohkuma, S. and B. Poole (1978). "Fluorescence probe measurement of the intralysosomal pH in living cells and the perturbation of pH by various agents." *Proc Natl Acad Sci U S A* 75(7): 3327-3331.
- Ohmori, H., E. C. Friedberg, *et al.* (2001). "The Y-family of DNA polymerases." *Mol Cell* 8(1): 7-8.
- Ohta, A., T. Obara, *et al.* (1985). "Molecular cloning of the *cls* gene responsible for cardiolipin synthesis in *Escherichia coli* and phenotypic consequences of its amplification." *J Bacteriol* 163(2): 506-514.
- Oksanen, J., R. Kindt, *et al.* (2007). "The vegan package." *Community ecology package*: 631-637.
- Omi, R., M. Goto, *et al.* (2004). "Expression, purification and preliminary X-ray characterization of histidinol phosphate phosphatase." *Acta Crystallogr D Biol Crystallogr* 60(Pt 3): 574-576.
- Oyeleke, S. B. and T. A. Okusanmi (2008). "Isolation and characterization of cellulose hydrolysing microorganism from the rumen of ruminants." *Afr J Biotechnol* 7(10).
- Ozment, C., J. Barchue, *et al.* (1999). "Structural study of *Escherichia coli* NAD synthetase: overexpression, purification, crystallization, and preliminary crystallographic analysis." *J Struct Biol* 127(3): 279-282.

- Padmanabha, J., J. Liu, *et al.* (2013). A methylotrophic methanogen isolate from the *Thermoplasmatales* affiliated RCC clade may provide insight into the role of this group in the rumen, In Proceedings of the 5th Greenhouse Gases and Animal Agriculture Conference, Dublin.
- Padyana, A. K. and S. K. Burley (2003). "Crystal structure of shikimate 5-dehydrogenase (SDH) bound to NADP: insights into function and evolution." *Structure* 11(8): 1005-1013.
- Paradis, E., K. Strimmer, *et al.* (2008). "The ape package." *Anal Phylogenet Evol.*
- Paris, S., P. M. Wessel, *et al.* (2002). "Overproduction, purification, and characterization of recombinant aspartate semialdehyde dehydrogenase from *Arabidopsis thaliana*." *Protein Express Purif* 24(1): 99-104.
- Patra, A. K. and Z. Yu (2012). "Effects of essential oils on methane production and fermentation by, and abundance and diversity of, rumen microbial populations." *Appl Environ Microbiol* 78(12): 4271-4280.
- Passarge, E., B. Horsthemke, *et al.* (1999). "Incorrect use of the term synteny." *Nat Genet* 23(4): 387.
- Paul, K., J. O. Nonoh, *et al.* (2012). "Methanoplasmatales," *Thermoplasmatales*-related archaea in termite guts and other environments, are the seventh order of methanogens." *Appl Environ Microbiol* 78(23): 8245-8253.
- Petersen, T. N., S. Brunak, *et al.* (2011). "SignalP 4.0: discriminating signal peptides from transmembrane regions." *Nat Methods* 8(10): 785-786.
- Petti, C. A. (2007). "Detection and identification of microorganisms by gene amplification and sequencing." *Clin Infect Dis* 44(8): 1108-1114.
- Pfaltz, A., A. Kobelt, *et al.* (1987). "Biosynthesis of coenzyme F<sub>430</sub> in methanogenic bacteria. Identification of 15,173-seco-F<sub>430</sub>-173-acid as an intermediate." *Eur J Biochem* 170(1-2): 459-467.
- Phillips, J. D., F. G. Whitby, *et al.* (2003). "Structural basis for tetrapyrrole coordination by uroporphyrinogen decarboxylase." *EMBO J* 22(23): 6225-6233.

- Phillipson, A. T. (1939). "The movements of the pouches of the stomach of sheep." *Q J Exp Physiol Cogn Med Sci* 29(4): 395-415.
- Pihl, T. D., S. Sharma, *et al.* (1994). "Growth phase-dependent transcription of the genes that encode the two methyl coenzyme M reductase isoenzymes and N<sup>5</sup>-methyltetrahydromethanopterin:coenzyme M methyltransferase in *Methanobacterium thermoautotrophicum* ΔH." *J Bacteriol* 176(20): 6384-6391.
- Pitson, S. M., G. L. Mendz, *et al.* (1999). "The tricarboxylic acid cycle of *Helicobacter pylori*." *Eur J Biochem* 260(1): 258-267.
- Pohlert, T. (2014). "The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR)." <https://cran.r-project.org/web/packages/PMCMR/index.html>
- Pollak, N., C. Dölle, *et al.* (2007). "The power to reduce: pyridine nucleotides - small molecules with a multitude of functions." *Biochem J* 402(2): 205-218.
- Poppi, D. P., D. J. Minson, *et al.* (1981). "Studies of cattle and sheep eating leaf and stem fractions of grasses. 1. The voluntary intake, digestibility and retention time in the reticulo-rumen." *Crop Pasture Sci* 32(1): 99-108.
- Porat, I., B. W. Waters, *et al.* (2004). "Two biosynthetic pathways for aromatic amino acids in the archaeon *Methanococcus maripaludis*." *J Bacteriol* 186(15): 4940-4950.
- Poulsen, M., C. Schwab, *et al.* (2013). "Methylophilic methanogenic *Thermoplasmata* implicated in reduced methane emissions from bovine rumen." *Nat Commun* 4: 1428.
- Prat, L., I. U. Heinemann, *et al.* (2012). "Carbon source-dependent expansion of the genetic code in bacteria." *Proc Natl Acad Sci U S A* 109(51): 21070-21075.
- Prieto, C., C. Garcia-Estrada, *et al.* (2012). "NRPSsp: non-ribosomal peptide synthase substrate predictor." *Bioinformatics* 28(3): 426-427.
- Primak, Y. A., M. Du, *et al.* (2011). "Characterization of a feedback-resistant mevalonate kinase from the archaeon *Methanosarcina mazei*." *Appl Environ Microbiol* 77(21): 7772-7778.
- Pugh, D. M. (2002). "The EU precautionary bans of animal feed additive antibiotics." *Toxicol Lett* 128(1-3): 35-44.
- Puigbo, P., I. G. Bravo, *et al.* (2008). "CAIcal: a combined set of tools to assess codon usage adaptation." *Biol Direct* 3: 38.

- Quitterer, F., P. Beck, *et al.* (2013). "Structure and reaction mechanism of pyrrolysine synthase (PylD)." *Angew Chem Int Ed Engl* 52(27): 7033-7037.
- Quitterer, F., A. List, *et al.* (2012). "Crystal structure of methylornithine synthase (PylB): insights into the pyrrolysine biosynthesis." *Angew Chem Int Ed Engl* 51(6): 1339-1342.
- Raffaelli, N., F. M. Pisani, *et al.* (1997). "Characterization of nicotinamide mononucleotide adenylyltransferase from thermophilic archaea." *J Bacteriol* 179(24): 7718-7723.
- Rahmstorf, S. and D. Coumou (2011). "Increase of extreme events in a warming world." *Proc Natl Acad Sci U S A* 108(44): 17905-17909.
- Randau, L., K. Calvin, *et al.* (2005). "The heteromeric *Nanoarchaeum equitans* splicing endonuclease cleaves noncanonical bulge-helix-bulge motifs of joined tRNA halves." *Proc Natl Acad Sci U S A* 102(50): 17934-17939.
- Rankin, C. A., G. C. Haslam, *et al.* (1993). "Sequence and expression of the gene for N<sup>10</sup>-formyltetrahydrofolate synthetase from *Clostridium cylindrosporium*." *Protein Sci* 2(2): 197-205.
- Rea, S., J. P. Bowman, *et al.* (2007). "*Methanobrevibacter millerae* sp. nov. and *Methanobrevibacter olleyae* sp. nov., methanogens from the ovine and bovine rumen that can utilize formate for growth." *Int J Syst Evol Microbiol* 57(3): 450-456.
- Rees, W. M. R., D. Lloyd, *et al.* (1995). "The effects of co-cultivation with the acetogen *Acetitomaculum ruminis* on the fermentative metabolism of the rumen fungi *Neocallimastix patriciarum* and *Neocallimastix* sp. strain L2." *FEMS Microbiol Lett* 133(1): 175-180.
- Reeve, J. N., J. Nolling, *et al.* (1997). "Methanogenesis: genes, genomes, and who's on first?" *J Bacteriol* 179(19): 5975-5986.
- Reguera, G., K. D. McCarthy, *et al.* (2005). "Extracellular electron transfer via microbial nanowires." *Nature* 435(7045): 1098-1101.
- Rhodes, D. and A. D. Hanson (1993). "Quaternary ammonium and tertiary sulfonium compounds in higher plants." *Annu Rev Plant Biol* 44(1): 357-384.
- Rice, P., I. Longden, *et al.* (2000). "EMBOSS: the European Molecular Biology Open Software Suite." *Trends Genet* 16(6): 276-277.

- Ricke, S. C., D. M. Schaefer, *et al.* (1994). "Influence of methylamine on anaerobic rumen bacterial growth and plant fiber digestion." *Bioresource Technol* 50(3): 253-257.
- Rinke, C., P. Schwientek, *et al.* (2013). "Insights into the phylogeny and coding potential of microbial dark matter." *Nature* 499(7459): 431-437.
- Ripley, B., B. Venables, *et al.* (2015). "Package 'MASS'."
- Ritchie, R. J. and J. Gibson (1987). "Permeability of ammonia and amines in *Rhodobacter sphaeroides* and *Bacillus firmus*." *Arch Biochem Biophys* 258(2): 332-341.
- Romantsov, T., Z. Guan, *et al.* (2009). "Cardiolipin and the osmotic stress responses of bacteria." *Biochim Biophys Acta* 1788(10): 2092-2100.
- Roongsawang, N., K. Washio, *et al.* (2010). "Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants." *Int J Mol Sci* 12(1): 141-172.
- Rosenthal, A. Z., E. G. Matson, *et al.* (2011). "RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture." *ISME J* 5(7): 1133-1142.
- Rospert, S., D. Linder, *et al.* (1990). "Two genetically distinct methyl-coenzyme M reductases in *Methanobacterium thermoautotrophicum* strain Marburg and  $\Delta H$ ." *Eur J Biochem* 194(3): 871-877.
- Rosewarne, C. P., P. Greenfield, *et al.* (Unpublished). "Draft genome sequence of *Methanobrevibacter acididurans* strain ATM<sup>T</sup> from anaerobic fermenter."
- Rotaru, A. E., M. P. Shrestha, *et al.* (2014). "A new model for electron flow during anaerobic digestion: direct interspecies electron transfer to *Methanosaeta* for the reduction of carbon dioxide to methane." *Energy Environ Sci* 7(1): 408-415.
- Rotaru, A. E., P. M. Shrestha, *et al.* (2014). "Direct interspecies electron transfer between *Geobacter metallireducens* and *Methanosarcina barkeri*." *Appl Environ Microbiol* 80(15): 4599-4605.
- Rowe, J. B., M. L. Loughnan, *et al.* (1979). "Secondary fermentation in the rumen of a sheep given a diet based on molasses." *Br J Nutr* 41(2): 393-397.
- Rstudio Team (2015). "Rstudio: integrated development for R. " Rstudio, Inc., Boston, MA

- Russell, J. B. and G. M. Cook (1995). "Energetics of bacterial growth: balance of anabolic and catabolic reactions." *Microbiol Rev* 59(1): 48-62.
- Rutherford, K., J. Parkhill, *et al.* (2000). "Artemis: sequence visualization and annotation." *Bioinformatics* 16(10): 944-945.
- Sachdeva, G., K. Kumar, *et al.* (2005). "SPAAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks." *Bioinformatics* 21(4): 483-491.
- Sakai, S., H. Imachi, *et al.* (2008). "*Methanocella paludicola* gen. nov., sp. nov., a methane-producing archaeon, the first isolate of the lineage 'Rice Cluster I', and proposal of the new archaeal order Methanocellales ord. nov." *Int J Syst Evol Microbiol* 58(Pt 4): 929-936.
- Sakamoto, N., A. M. Kotre, *et al.* (1975). "Glutamate dehydrogenase from *Escherichia coli*: purification and properties." *J Bacteriol* 124(2): 775-783.
- Saleem, F., S. Bouatra, *et al.* (2013). "The bovine ruminal fluid metabolome." *Metabolomics* 9(2): 360-378.
- Samuel, B. S., E. E. Hansen, *et al.* (2007). "Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut." *Proc Natl Acad Sci U S A* 104(25): 10643-10648.
- Samuel, G. and P. Reeves (2003). "Biosynthesis of *O*-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and *O*-antigen assembly." *Carbohydr Res* 338(23): 2503-2519.
- Sandkvist, M. (2001). "Biology of type II secretion." *Mol Microbiol* 40(2): 271-283.
- Santiveri, C. M. and M. A. Jimenez (2010). "Tryptophan residues: scarce in proteins but strong stabilizers of  $\beta$ -hairpin peptides." *Biopolymers* 94(6): 779-790.
- Sasarman, A., A. Nepveu, *et al.* (1987). "Molecular cloning and sequencing of the *hemD* gene of *Escherichia coli* K-12 and preliminary data on the Uro operon." *J Bacteriol* 169(9): 4257-4262.
- Sato, T., H. Atomi, *et al.* (2007). "Archaeal type III RuBisCOs function in a pathway for AMP metabolism." *Science* 315(5814): 1003-1006.

- Sauer, K. and R. K. Thauer (1998). "Methanol:coenzyme M methyltransferase from *Methanosarcina barkeri*--identification of the active-site histidine in the corrinoid-harboring subunit MtaC by site-directed mutagenesis." *Eur J Biochem* 253(3): 698-705.
- Saunders, M. A. and A. S. Lea (2008). "Large contribution of sea surface warming to recent increase in Atlantic hurricane activity." *Nature* 451(7178): 557-560.
- Savant, D. V., Y. S. Shouche, *et al.* (2002). "*Methanobrevibacter acididurans* sp. nov., a novel methanogen from a sour anaerobic digester." *Int J Syst Evol Microbiol* 52(4): 1081-1087.
- Sayers, E. W., T. Barrett, *et al.* (2010). "Database resources of the National Center for Biotechnology Information." *Nucleic Acids Res* 38(Database issue): D5-16.
- Scanlan, P. D., F. Shanahan, *et al.* (2008). "Human methanogen diversity and incidence in healthy and diseased colonic groups using *mcrA* gene analysis." *BMC Microbiol* 8: 79.
- Schäfer, G., M. Engelhard, *et al.* (1999). "Bioenergetics of the Archaea." *Microbiol Mol Biol Rev* 63(3): 570-620.
- Scheller, S., M. Goenrich, *et al.* (2010). "The key nickel enzyme of methanogenesis catalyses the anaerobic oxidation of methane." *Nature* 465(7298): 606-608.
- Schiermeier, Q. (2011). "Increased flood risk linked to global warming." *Nature* 470(7334): 316.
- Schlaberg, R., K. E. Simmon, *et al.* (2012). "A systematic approach for discovering novel, clinically relevant bacteria." *Emerg Infect Dis* 18(3): 422-430.
- Schlame, M. (2008). "Thematic Review Series: Glycerolipids. Cardiolipin synthesis for the assembly of bacterial and mitochondrial membranes." *J Lipid Res* 49(8): 1607-1620.
- Schlame, M. and M. Ren (2009). "The role of cardiolipin in the structural organization of mitochondrial membranes." *Biochim Biophys Acta* 1788(10): 2080-2083.
- Schleifer, K. H. and O. Kandler (1972). "Peptidoglycan types of bacterial cell walls and their taxonomic implications." *Bacteriol Rev* 36(4): 407-477.
- Schluter, D., E. A. Clifford, *et al.* (2004). "Parallel evolution and inheritance of quantitative traits." *Am Nat* 163(6): 809-822.



- Schultz, J. E. and A. Matin (1991). "Molecular and functional characterization of a carbon starvation gene of *Escherichia coli*." J Mol Biol 218(1): 129-140.
- Scofield, M. A., W. S. Lewis, *et al.* (1990). "Nucleotide sequence of *Escherichia coli asnB* and deduced amino acid sequence of asparagine synthetase B." J Biol Chem 265(22): 12895-12902.
- Seedorf, H., S. Kittelmann, *et al.* (2014). "RIM-DB: a taxonomic framework for community structure analysis of methanogenic archaea from the rumen and other intestinal environments." PeerJ 2: e494.
- Seedorf, H., S. Kittelmann, *et al.* (2015). "Few highly abundant operational taxonomic units dominate within rumen methanogenic archaeal species in New Zealand sheep and cattle." Appl Environ Microbiol 81(3): 986-995.
- Sekowska, A., P. Bertin, *et al.* (1998). "Characterization of polyamine synthesis pathway in *Bacillus subtilis* 168." Mol Microbiol 29(3): 851-858.
- Selbmann, L., D. Isola, *et al.* (2012). "Potential extinction of Antarctic endemic fungal species as a consequence of global warming." Sci Total Environ 438: 127-134.
- Serina, L., C. Blondin, *et al.* (1995). "*Escherichia coli* UMP-kinase, a member of the aspartokinase family, is a hexamer regulated by guanine nucleotides and UTP." Biochemistry 34(15): 5066-5074.
- Setzke, E., R. Hedderich, *et al.* (1994). "H<sub>2</sub>: heterodisulfide oxidoreductase complex from *Methanobacterium thermoautotrophicum*. Composition and properties." Eur J Biochem 220(1): 139-148.
- Shariat, N., R. E. Timme, *et al.* (2015). "Characterization and evolution of *Salmonella* CRISPR-Cas systems." Microbiology 161: 374-386.
- Sharp, P. M. and W. H. Li (1987). "The codon adaptation index--a measure of directional synonymous codon usage bias, and its potential applications." Nucleic Acids Res 15(3): 1281-1295.
- She, Q., R. K. Singh, *et al.* (2001). "The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2." PNAS 98(14): 7835-7840.

- Shen, B. W., D. H. Dyer, *et al.* (1999). "The crystal structure of a bacterial, bifunctional 5, 10 methylene-tetrahydrofolate dehydrogenase/cyclohydrolase." *Prot Sci* 8(6): 1342-1349.
- Shrestha, P. M., A. E. Rotaru, *et al.* (2013). "Transcriptomic and genetic analysis of direct interspecies electron transfer." *Appl Environ Microbiol* 79(7): 2397-2404.
- Shi, W., C. D. Moon, *et al.* (2014). "Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome." *Genome Res* 24(9): 1517-1525.
- Shieh, J. S. and W. B. Whitman (1987). "Pathway of acetate assimilation in autotrophic and heterotrophic methanococci." *J Bacteriol* 169(11): 5327-5329.
- Shima, S., E. Warkentin, *et al.* (2002). "Structure and function of enzymes involved in the methanogenic pathway utilizing carbon dioxide and molecular hydrogen." *J Biosci Bioeng* 93(6): 519-530.
- Shimizu, H., S. Yamagata, *et al.* (2001). "Cloning and overexpression of the *oah1* gene encoding *O*-acetyl-L-homoserine sulfhydrylase of *Thermus thermophilus* HB8 and characterization of the gene product." *Biochim Biophys Acta* 1549(1): 61-72.
- Shimon, L. J., E. A. Bayer, *et al.* (1997). "A cohesin domain from *Clostridium thermocellum*: the crystal structure provides new insights into cellulosome assembly." *Structure* 5(3): 381-390.
- Shin, E. C., B. R. Choi, *et al.* (2004). "Phylogenetic analysis of archaea in three fractions of cow rumen based on the 16S rDNA sequence." *Anaerobe* 10(6): 313-319.
- Shinkai, T., O. Enishi, *et al.* (2012). "Mitigation of methane production from cattle by feeding cashew nut shell liquid." *J Dairy Sci* 95(9): 5308-5316.
- Shinzato, N., T. Matsumoto, *et al.* (1999). "Phylogenetic diversity of symbiotic methanogens living in the hindgut of the lower termite *Reticulitermes speratus* analyzed by PCR and *in situ* hybridization." *Appl Environ Microbiol* 65(2): 837-840.
- Siguier, P., E. Gourbeyre, *et al.* (2015). "Everyman's guide to bacterial insertion sequences." *Microbiol Spectr* 3(2): MDNA3-0030-2014.
- Siguier, P., J. Perochon, *et al.* (2006). "ISfinder: the reference centre for bacterial insertion sequences." *Nucleic Acids Res* 34(Database issue): D32-36.

- Silk, G. W., B. F. Matthews, *et al.* (1994). "Cloning and expression of the soybean *dapA* gene encoding dihydrodipicolinate synthase." *Plant Mol Biol* 26(3): 989-993.
- Skillman, L. C., P. N. Evans, *et al.* (2004). "16S ribosomal DNA-directed PCR primers for ruminal methanogens and identification of methanogens colonising young lambs." *Anaerobe* 10(5): 277-285.
- Silva, K. L., E. R. Duarte, *et al.* (2014). "Protozoários ruminais de novilhos de corte criados em pastagem tropical durante o período seco." *Ciênc Anim Bras* 15: 259-265.
- Simcock, D. C., K. N. Joblin, *et al.* (1999). "Hypergastrinaemia, abomasal bacterial population densities and pH in sheep infected with *Ostertagia circumcincta*." *Int J Parasitol* 29(7): 1053-1063.
- Siragusa, R. J., J. J. Cerda, *et al.* (1988). "Methanol production from the degradation of pectin by human colonic bacteria." *Am J Clin Nutr* 47(5): 848-851.
- Smith, D. R., L. A. Doucette-Stamm, *et al.* (1997). "Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: functional analysis and comparative genomics." *J Bacteriol* 179(22): 7135-7155.
- Smith, E. A. and G. T. Macfarlane (1996). "Studies on amine production in the human colon: Enumeration of amine forming bacteria and physiological effects of carbohydrate and pH." *Anaerobe* 2(5): 285-297.
- Smith, J. M. and H. A. Daum, 3rd (1986). "Nucleotide sequence of the *purM* gene encoding 5'-phosphoribosyl-5-aminoimidazole synthetase of *Escherichia coli* K12." *J Biol Chem* 261(23): 10632-10636.
- Smith, K. S. and J. G. Ferry (1999). "A plant-type (beta-class) carbonic anhydrase in the thermophilic methanoarchaeon *Methanobacterium thermoautotrophicum*." *J Bacteriol* 181(20): 6247-6253.
- Smith, P. H. and R. E. Hungate (1958). "Isolation and characterization of *Methanobacterium ruminantium* n. sp." *Journal of Bacteriology* 75(6): 713-718.
- Smith, T. F. and M. S. Waterman (1981). "Identification of common molecular subsequences." *J Mol Biol* 147(1): 195-197.

- Sneath, P. H. and R. R. Sokal (1962). "Numerical taxonomy." *Nature* 193: 855-860.
- Soliva, C. R., I. K. Hindrichsen, *et al.* (2003). "Effects of mixtures of lauric and myristic acid on rumen methanogens and methanogenesis *in vitro*." *Lett Appl Microbiol* 37(1): 35-39.
- Sollinger, A., C. Schwab, *et al.* (2016). "Phylogenetic and genomic analysis of Methanomassiliicoccales in wetlands and animal intestinal tracts reveals clade-specific habitat preferences." *FEMS Microbiol Ecol* 92(1).
- Soupene, E., L. He, *et al.* (1998). "Ammonia acquisition in enteric bacteria: physiological role of the ammonium/methylammonium transport B (AmtB) protein." *Proc Natl Acad Sci U S A* 95(12): 7030-7034.
- Sparrow, C. P. and C. R. Raetz (1985). "Purification and properties of the membrane-bound CDP-diglyceride synthetase from *Escherichia coli*." *J Biol Chem* 260(22): 12084-12091.
- Sprenger, G. A. (1995). "Genetics of pentose-phosphate pathway enzymes of *Escherichia coli* K-12." *Arch Microbiol* 164(5): 324-330.
- Sprenger, G. A., U. Schorken, *et al.* (1995). "Transketolase A of *Escherichia coli* K12. Purification and properties of the enzyme from recombinant strains." *Eur J Biochem* 230(2): 525-532.
- Sprenger, G. A., U. Schörken, *et al.* (1995). "Transaldolase B of *Escherichia coli* K-12: cloning of its gene, *talB*, and characterization of the enzyme from recombinant strains." *J Bacteriol* 177(20): 5930-5936.
- Sprott, G. D., I. Ekiel, *et al.* (1993). "Metabolic pathways in *Methanococcus jannaschii* and other methanogenic bacteria." *Appl Environ Microbiol* 59(4): 1092-1098.
- Srinivasan, G., C. M. James, *et al.* (2002). "Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA." *Science* 296(5572): 1459-1462.
- Stackebrandt, E. and J. Ebers (2006). "Taxonomic parameters revisited: tarnished gold standards." *Microbiol today* 33(4): 152.
- Staden, R., K. F. Beal, *et al.* (2000). "The Staden package, 1998." *Methods Mol Biol* 132: 115-130.

- Staples, C. R., S. Lahiri, *et al.* (2007). "Expression and association of group IV nitrogenase NifD and NifH homologs in the non-nitrogen-fixing archaeon *Methanocaldococcus jannaschii*." J Bacteriol 189(20): 7392-7398.
- Staudenmaier, H., B. Van Hove, *et al.* (1989). "Nucleotide sequences of the *fecBCDE* genes and locations of the proteins suggest a periplasmic-binding-protein-dependent transport mechanism for iron(III) dicitrate in *Escherichia coli*." J Bacteriol 171(5): 2626-2633.
- Steenhoudt, O. and J. Vanderleyden (2000). "*Azospirillum*, a free-living nitrogen-fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects." FEMS Microbiol Rev 24(4): 487-506.
- Stotz, A. and P. Linder (1990). "The ADE2 gene from *Saccharomyces cerevisiae*: sequence and new vectors." Gene 95(1): 91-98.
- Strieker, M., A. Tanovic, *et al.* (2010). "Nonribosomal peptide synthetases: structures and dynamics." Curr Opin Struct Biol 20(2): 234-240.
- Stupperich, E., K. E. Hammel, *et al.* (1983). "Carbon monoxide fixation into the carboxyl group of acetyl coenzyme A during autotrophic growth of *Methanobacterium*." FEBS Lett 152(1): 21-23.
- Subharat, S., D. Shu, *et al.* (2015). "Vaccination of cattle with a methanogen protein produces specific antibodies in the saliva which are stable in the rumen." Vet Immunol Immunopathol 164(3-4): 201-207.
- Tajima, K., T. Nagamine, *et al.* (2001). "Phylogenetic analysis of archaeal 16S rRNA libraries from the rumen suggests the existence of a novel group of archaea not associated with known methanogens." FEMS Microbiol Lett 200(1): 67-72.
- Takai, K. and K. Horikoshi (1999). "Genetic diversity of archaea in deep-sea hydrothermal vent environments." Genetics 152(4): 1285-1297.
- Tallant, T. C. and J. A. Krzycki (1997). "Methylthiol:coenzyme M methyltransferase from *Methanosarcina barkeri*, an enzyme of methanogenesis from dimethylsulfide and methylmercaptopropionate." J Bacteriol 179(22): 6902-6911.

- Tallant, T. C., L. Paul, *et al.* (2001). "The MtsA subunit of the methylthiol:coenzyme M methyltransferase of *Methanosarcina barkeri* catalyses both half-reactions of corrinoid-dependent dimethylsulfide: coenzyme M methyl transfer." *J Biol Chem* 276(6): 4485-4493.
- Tamura, K., G. Stecher, *et al.* (2013). "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0." *Mol Biol Evol* 30(12): 2725-2729.
- Tan, H. Y., C. C. Sieo, *et al.* (2011). "Effects of condensed tannins from *Leucaena* on methane production, rumen fermentation and populations of methanogens and protozoa *in vitro*." *Anim Feed Sci Technol* 169(3-4): 185-193.
- Tan, H. Y., C. C. Sieo, *et al.* (2011). "Diversity of bovine rumen methanogens *In vitro* in the presence of condensed tannins, as determined by sequence analysis of 16S rRNA gene library." *J Microbiol* 49(3): 492-498.
- Tang, X., S. Ezaki, *et al.* (1999). "The tryptophan biosynthesis gene cluster *trpCDEGFBA* from *Pyrococcus kodakaraensis* KOD1 is regulated at the transcriptional level and expressed as a single mRNA." *MGG* 262(4-5): 815-821.
- Tatusov, R. L., D. A. Natale, *et al.* (2001). "The COG database: new developments in phylogenetic classification of proteins from complete genomes." *Nucleic Acids Res* 29(1): 22-28.
- Tavendale, M. H., L. P. Meagher, *et al.* (2005). "Methane production from *in vitro* rumen incubations with *Lotus pedunculatus* and *Medicago sativa*, and effects of extractable condensed tannin fractions on methanogenesis." *Anim Feed Sci Technol* 123-124 Part 1: 403-419.
- Taylor, C. D., B. C. McBride, *et al.* (1974). "Coenzyme M, essential for growth of a rumen strain of *Methanobacterium ruminantium*." *J Bacteriol* 120(2): 974-975.
- Taylor, C. D. and R. S. Wolfe (1974). "Structure and methylation of coenzyme M(HSCH<sub>2</sub>CH<sub>2</sub>SO<sub>3</sub>)." *J Biol Chem* 249(15): 4879-4885.
- Teeling, H., J. Waldmann, *et al.* (2004). "TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences." *BMC Bioinformatics* 5: 163.
- Temin, H. M. (1981). "Structure, variation and synthesis of retrovirus long terminal repeat." *Cell* 27(1 Pt 2): 1-3.

- Temin, H. M. (1981). "Structure, variation and synthesis of retrovirus long terminal repeat." *Cell* 27(1 Pt 2): 1-3.
- Tesmer, J. J., T. L. Stemmler, *et al.* (1994). "Preliminary X-ray analysis of *Escherichia coli* GMP synthetase: determination of anomalous scattering factors for a cysteinyl mercury derivative." *Proteins* 18(4): 394-403.
- Thauer, R. K. and L. G. Bonacker (1994). "Biosynthesis of coenzyme F<sub>430</sub>, a nickel porphinoid involved in methanogenesis." *Ciba Found Symp* 180: 210-222.
- Thauer, R. K. (1998). "Biochemistry of methanogenesis: a tribute to Marjory Stephenson. 1998 Marjory Stephenson Prize Lecture." *Microbiology* 144 ( Pt 9): 2377-2406.
- Thauer, R. K., A. K. Kaster, *et al.* (2008). "Methanogenic archaea: ecologically relevant differences in energy conservation." *Nat Rev Microbiol* 6: 579-591.
- Thauer, R. K., A. K. Kaster, *et al.* (2010). "Hydrogenases from methanogenic archaea, nickel, a novel cofactor, and H<sub>2</sub> storage." *Annu Rev Biochem* 79: 507-536.
- Thoden, J. B., G. N. Phillips, Jr., *et al.* (2001). "Molecular structure of dihydroorotase: a paradigm for catalysis through the use of a binuclear metal center." *Biochemistry* 40(24): 6989-6997.
- Thoden, J. B., F. M. Raushel, *et al.* (1999). "The structure of carbamoyl phosphate synthetase determined to 2.1 Å resolution." *Acta Crystallogr D Biol Crystallogr* 55(Pt 1): 8-24.
- Thompson, G. A., Jr. (1969). "The metabolism of 2-aminoethylphosphonate lipids in *Tetrahymena pyriformis*." *Biochim Biophys Acta* 176(2): 330-338.
- Thompson, J. D., T. J. Gibson, *et al.* (2002). "Multiple sequence alignment using ClustalW and ClustalX." *Curr Protoc Bioinformatics* Chapter 2: Unit 2 3.
- Thompson, L. D., L. D. Brandon, *et al.* (1989). "Transfer RNA intron processing in the halophilic archaeobacteria." *Can J Microbiol* 35(1): 36-42.
- Tietze, M., A. Beuchle, *et al.* (2003). "Redox potentials of methanophenazine and CoB-S-S-CoM, factors involved in electron transport in Methanogenic archaea." *Chembiochem* 4(4): 333-335.

- Tock, M. R. and D. T. Dryden (2005). "The biology of restriction and anti-restriction." *Curr Opin Microbiol* 8(4): 466-472.
- Ton-That, H., L. A. Marraffini, *et al.* (2004). "Protein sorting to the cell wall envelope of Gram-positive bacteria." *Biochim Biophys Acta* 1694(1-3): 269-278.
- Torrents, E., G. Buist, *et al.* (2000). "The anaerobic (class III) ribonucleotide reductase from *Lactococcus lactis*. Catalytic properties and allosteric regulation of the pure enzyme system." *J Biol Chem* 275(4): 2463-2471.
- Towne, G., T. G. Nagaraja, *et al.* (1990). "Dynamics of ruminal ciliated protozoa in feedlot cattle." *Appl Environ Microbiol* 56(10): 3174-3178.
- Tschech, A. and N. Pfennig (1984). "Growth yield increase linked to caffeate reduction in *Acetobacterium woodii*." *Arch Microbiol* 137(2): 163-167.
- Tsukagoshi, N., G. Tamura, *et al.* (1970). "A novel protoplast-bursting factor (surfactin) obtained from *Bacillus subtilis* IAM 1213. I. The effects of surfactin on *Bacillus megaterium* KM." *Biochim Biophys Acta* 196(2): 204-210.
- Tu, J. and W. Zillig (1982). "Organization of rRNA structural genes in the archaeobacterium *Thermoplasma acidophilum*." *Nucleic Acids Res* 10(22).
- Tukey, J. W. (1949). "Comparing individual means in the analysis of variance." *Biometrics* 5(2): 99-114.
- Uchiyama, T., K. Ito, *et al.* (2010). "Iron-corroding methanogen isolated from a crude-oil storage tank." *Appl Environ Microbiol* 76(6): 1783-1788.
- Van Soest, P. J. (1984). "Some physical characteristics of dietary fibres and their influence on the microbial ecology of the human colon." *Proc Nutr Soc* 43(1): 25-33.
- Vantcheva, Z. M., K. Prodhan, *et al.* (1970). "Rumen methanol *in vivo* and *in vitro*." *J Dairy Sci* 53(10): 1511-1514.
- Vartak, N. B., L. Liu, *et al.* (1991). "A functional *leuABCD* operon is required for leucine synthesis by the tyrosine-repressible transaminase in *Escherichia coli* K-12." *J Bacteriol* 173(12): 3864-3871.



- Veiga, M. C., M. K. Jain, *et al.* (1997). "Composition and role of extracellular polymers in methanogenic granules." *Appl Environ Microbiol* 63(2): 403-407.
- Vernikos, G. S. and J. Parkhill (2006). "Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands." *Bioinformatics* 22(18): 2196-2203.
- Vesth, T., K. Lagesen, *et al.* (2013). "CMG-Biotools, a free workbench for basic comparative microbial genomics." *PloS one* 8(4): e60120.
- Vinokur, J. M., T. P. Korman, *et al.* (2014). "Evidence of a novel mevalonate pathway in archaea." *Biochemistry* 53(25): 4161-4168.
- Wadhwa, D. R. and A. D. Care (2002). "The absorption of phosphate ions from the ovine reticulorumen." *Vet J* 163(2): 182-186.
- Wallace, R. J., N. R. McEwan, *et al.* (2002). "Natural products as manipulators of rumen fermentation." *Asian Australas. J. Anim. Sci* 15(10): 1458-1468.
- Wang, H., D. P. Fewer, *et al.* (2014). "Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes." *Proc Natl Acad Sci U S A* 111(25): 9259-9264.
- Wang, J., K. A. Stieglitz, *et al.* (2005). "Structural basis for ordered substrate binding and cooperativity in aspartate transcarbamoylase." *Proc Natl Acad Sci U S A* 102(25): 8881-8886.
- Wang, L., S. Chen, *et al.* (2007). "Phosphorothioation of DNA in bacteria by *dnd* genes." *Nat Chem Biol* 3(11): 709-710.
- Ward, D. E., S. W. Kengen, *et al.* (2000). "Purification and characterization of the alanine aminotransferase from the hyperthermophilic Archaeon *Pyrococcus furiosus* and its role in alanine production." *J Bacteriol* 182(9): 2559-2566.
- Wargo, M. J. (2013). "Homeostasis and catabolism of choline and glycine betaine: lessons from *Pseudomonas aeruginosa*." *Appl Environ Microbiol* 79(7): 2112-2120.
- Watanabe, T., S. Asakawa, *et al.* (2004). "DGGE method for analyzing 16S rDNA of methanogenic archaeal community in paddy field soil." *FEMS Microbiol Lett* 232(2): 153-163.

- Watkins, A. J., E. G. Roussel, *et al.* (2014). "Glycine betaine as a direct substrate for methanogens (*Methanococcoides* spp.)." *Appl Environ Microbiol* 80(1): 289-293.
- Watson, G. M., J. P. Yu, *et al.* (1999). "Unusual ribulose 1,5-bisphosphate carboxylase/oxygenase of anoxic Archaea." *J Bacteriol* 181(5): 1569-1575.
- Wassenaar, R. W., P. J. Daas, *et al.* (1996). "Involvement of methyltransferase-activating protein and methyltransferase 2 isoenzyme II in methylamine:coenzyme M methyltransferase reactions in *Methanosarcina barkeri* Fusaro." *J Bacteriol* 178(23): 6937-6944.
- Wegener, G., V. Krukenberg, *et al.* (2015). "Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria." *Nature* 526(7574): 587-590.
- Weidner, S., A. Becker, *et al.* (2012). "Genome sequence of the soybean symbiont *Sinorhizobium fredii* HH103." *J Bacteriol* 194(6): 1617-1618.
- Weisburg, W. G., S. M. Barns, *et al.* (1991). "16S ribosomal DNA amplification for phylogenetic study." *J Bacteriol* 173(2): 697-703.
- Welander, P. V. and W. W. Metcalf (2005). "Loss of the mtr operon in *Methanosarcina* blocks growth on methanol, but not methanogenesis, and reveals an unknown methanogenic pathway." *Proc Natl Acad Sci U S A* 102(30): 10664-10669.
- Weljie, A. M., J. Newton, *et al.* (2006). "Targeted profiling: quantitative analysis of 1H NMR metabolomics data." *Anal Chem* 78(13): 4430-4442.
- Welte, C. and U. Deppenmeier (2011). "Membrane-bound electron transport in *Methanosaeta thermophila*." *J Bacteriol* 193(11): 2868-2870.
- Welte, C. and U. Deppenmeier (2011). "Re-evaluation of the function of the F<sub>420</sub> dehydrogenase in electron transport of *Methanosarcina mazei*." *FEBS J* 278(8): 1277-1287.
- Welte, C. and U. Deppenmeier (2014). "Bioenergetics and anaerobic respiratory chains of acetoclastic methanogens." *Biochim Biophys Acta* 1837(7): 1130-1147.
- White, P. J., G. Millar, *et al.* (1988). "The overexpression, purification and complete amino acid sequence of chorismate synthase from *Escherichia coli* K12 and its comparison with the enzyme from *Neurospora crassa*." *Biochem. J* 251: 313-322.

- White, R. H. (1985). "Biosynthesis of coenzyme M (2-mercaptoethanesulfonic acid)." *Biochemistry* 24(23): 6487-6493.
- White, R. H. (2003). "The biosynthesis of cysteine and homocysteine in *Methanococcus jannaschii*." *BBA-Gen Subjects* 1624(1): 46-53.
- White, R. H. (2004). "L-Aspartate semialdehyde and a 6-deoxy-5-ketohexose 1-phosphate are the precursors to the aromatic amino acids in *Methanocaldococcus jannaschii*." *Biochemistry* 43(23): 7618-7627.
- White, R. H. and H. Xu (2006). "Methylglyoxal is an intermediate in the biosynthesis of 6-deoxy-5-ketofructose-1-phosphate: a precursor for aromatic amino acid biosynthesis in *Methanocaldococcus jannaschii*." *Biochemistry* 45(40): 12366-12379.
- White, R. W. (1969). "Viable bacteria inside the rumen ciliate *Entodinium caudatum*." *J Gen Microbiol* 56(3): 403-408.
- Widdel, F. and N. Pfennig (1981). "Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. I. Isolation of new sulfate-reducing bacteria enriched with acetate from saline environments. Description of *Desulfobacter postgatei* gen. nov., sp. nov." *Arch Microbiol* 129(5): 395-400.
- Widyastuti, Y., S. K. Lee, *et al.* (1992). "Isolation and characterization of rice-straw degrading *Clostridia* from cattle rumen." *J Vet Med Sci* 54(1): 185-188.
- Wiedemann, I., T. Bottiger, *et al.* (2006). "The mode of action of the lantibiotic lacticin 3147--a complex mechanism involving specific interaction of two peptides and the cell wall precursor lipid II." *Mol Microbiol* 61(2): 285-296.
- Willcott, M. R. (2009). "MestRe Nova." *J Am Chem Soc* 131(36): 13180-13180.
- Wilding, E. I., J. R. Brown, *et al.* (2000). "Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in Gram-positive cocci." *J Bacteriol* 182(15): 4319-4327.
- Wishart, D. S., T. Jewison, *et al.* (2013). "HMDB 3.0--The Human Metabolome Database in 2013." *Nucleic Acids Res* 41(Database issue): D801-807.

- Witting, K. and R. D. Sussmuth (2011). "Discovery of antibacterials and other bioactive compounds from microorganisms-evaluating methodologies for discovery and generation of non-ribosomal peptide antibiotics." *Curr Drug Targets* 12(11): 1547-1559.
- Wolfram, F., E. N. Kitova, *et al.* (2014). "Catalytic mechanism and mode of action of the periplasmic alginate epimerase AlgG." *J Biol Chem* 289(9): 6006-6019.
- Wolin, M. J. (1976). Interactions between H<sub>2</sub>-producing and methane-producing species. Microbial production and utilization of gases (H<sub>2</sub>, CH<sub>4</sub>, CO). H. G. Schlegel, E. G. Gottingen and N. Pfennig: 141-150.
- Wolin, M. J., T. L. Miller, *et al.* (1997). Microbe-microbe interactions. The rumen microbial ecosystem, Springer: 467-491.
- Wong, T. Y. and S. D. Schwartzbach (2015). "Protein mis-termination initiates genetic diseases, cancers, and restricts bacterial genome expansion." *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev* 33(3): 255-285.
- Wood, G. E., A. K. Haydock, *et al.* (2003). "Function and regulation of the formate dehydrogenase genes of the methanogenic archaeon *Methanococcus maripaludis*." *J Bacteriol* 185(8): 2548-2554.
- Wright, A. D., A. F. Toovey, *et al.* (2006). "Molecular identification of methanogenic archaea from sheep in Queensland, Australia reveal more uncultured novel archaea." *Anaerobe* 12(3): 134-139.
- Wright, A. D. G., C. H. Auckland, *et al.* (2007). "Molecular diversity of methanogens in feedlot cattle from Ontario and Prince Edward Island, Canada." *Appl Environ Microbiol* 73(13): 4206-4210.
- Wubbolts, M. G., P. Terpstra, *et al.* (1990). "Variation of cofactor levels in *Escherichia coli*. Sequence analysis and expression of the *pncB* gene encoding nicotinic acid phosphoribosyltransferase." *J Biol Chem* 265(29): 17665-17672.
- Wugeditsch, T., N. E. Zachara, *et al.* (1999). "Structural heterogeneity in the core oligosaccharide of the S-layer glycoprotein from *Aneurinibacillus thermoaerophilus* DSM 10155." *Glycobiology* 9(8): 787-795.

www.stats.govt.nz. (2015). "Global New Zealand – International trade, investment, and travel profile: Year ended December 2014." from www.stats.govt.nz.

Xia, J., R. Mandal, *et al.* (2012). "MetaboAnalyst 2.0--a comprehensive server for metabolomic data analysis." *Nucleic Acids Res* 40(Web Server issue): W127-133.

Xia, J., I. V. Sinelnikov, *et al.* (2015). "MetaboAnalyst 3.0-making metabolomics more meaningful." *Nucleic Acids Res* 43(W1): W251-257.

Xing, R. Y. and W. B. Whitman (1991). "Characterization of enzymes of the branched-chain amino acid biosynthetic pathway in *Methanococcus* spp." *J Bacteriol* 173(6): 2086-2092.

Yamagata, S. (1987). "Partial purification and some properties of homoserine *O*-acetyltransferase of a methionine auxotroph of *Saccharomyces cerevisiae*." *J Bacteriol* 169(8): 3458-3463.

Yamagata, S., M. Isaji, *et al.* (1994). "Overexpression of the *Saccharomyces cerevisiae* MET17/MET25 gene in *Escherichia coli* and comparative characterization of the product with *O*-acetylserine: *O*-acetylhomoserine sulfhydrylase of the yeast." *Appl Microbiol Biotechnol* 42(1): 92-99.

Yamagishi, T., S. Ishida, *et al.* (1964). "Isolation of toxigenic strains of *Clostridium perfringens* from soil." *J Bacteriol* 88: 646-652.

Yamao, F., A. Muto, *et al.* (1985). "UGA is read as tryptophan in *Mycoplasma capricolum*." *Proc Natl Acad Sci U S A* 82(8): 2306-2309.

Yamashita, S., H. Hemmi, *et al.* (2004). "Type 2 isopentenyl diphosphate isomerase from a thermoacidophilic archaeon *Sulfolobus shibatae*." *Eur J Biochem* 271(6): 1087-1093.

Yanagita, K., Y. Kamagata, *et al.* (2000). "Phylogenetic analysis of methanogens in sheep rumen ecosystem and detection of *Methanomicrobium mobile* by fluorescence *in situ* hybridization." *Biosci Biotechnol Biochem* 64(8): 1737-1742.

Yang, Z., A. Savchenko, *et al.* (2003). "Aspartate dehydrogenase, a novel enzyme identified from structural and functional studies of TM1643." *J Biol Chem* 278(10): 8804-8808.

Yasmin, A., J. G. Kenny, *et al.* (2010). "Comparative genomics and transduction potential of *Enterococcus faecalis* temperate bacteriophages." *J Bacteriol* 192(4): 1122-1130.

- Yasueda, H., Y. Kawahara, *et al.* (1999). "Bacillus subtilis *yckG* and *yckF* encode two key enzymes of the ribulose monophosphate pathway used by methylophs, and *yckH* is required for their expression." J Bacteriol 181(23): 7154-7160.
- Yip, W. S., N. G. Vincent, *et al.* (2013). "Ribonucleoproteins in archaeal pre-rRNA processing and modification." Archaea 2013: 614735.
- Yoshihisa, T. (2014). "Handling tRNA introns, archaeal way and eukaryotic way." Front Genet 5: 213.
- Yoshinaga, M. Y., L. Wormer, *et al.* (2012). "Novel cardiolipins from uncultured methane-metabolizing archaea." Archaea 2012: 832097.
- Yoshioka, Y., S. Kurei, *et al.* (2001). "Identification of a monofunctional aspartate kinase gene of *Arabidopsis thaliana* with spatially and temporally regulated expression." Genes Genet Syst 76(3): 189-198.
- Yu, N. Y., J. R. Wagner, *et al.* (2010). "PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes." Bioinformatics 26(13): 1608-1615.
- Zadoks, R. N., H. M. Griffiths, *et al.* (2011). "Sources of *Klebsiella* and *Raoultella* species on dairy farms: be careful where you walk." J Dairy Sci 94(2): 1045-1051.
- Zamenhof, S. and H. H. Eichhorn (1967). "Study of microbial evolution through loss of biosynthetic functions: establishment of "defective" mutants." Nature 216(5114): 456-45
- Zeikus, J. G. and D. L. Henning (1975). "*Methanobacterium arbophilicum* sp.nov. An obligate anaerobe isolated from wetwood of living trees." Antonie Van Leeuwenhoek 41(4): 543-552.
- Zerbino, D. R. (2010). "Using the Velvet de novo assembler for short-read sequencing technologies." Curr Protoc Bioinformatics Chapter 11: Unit 11 15.
- Zhang, D. L., L. Daniels, *et al.* (1990). "Biosynthesis of archaebacterial membranes. Formation of isoprene ethers by a prenyl transfer reaction." J Am Chem Soc 112(3): 1264-1265.
- Zhang, J. K., A. K. White, *et al.* (2002). "Directed mutagenesis and plasmid-based complementation in the methanogenic archaeon *Methanosarcina acetivorans* C2A demonstrated by genetic analysis of proline biosynthesis." J Bacteriol 184(5): 1449-1454.

Zhang, Y., P. V. Baranov, *et al.* (2005). "Pyrrolysine and selenocysteine use dissimilar decoding strategies." *J Biol Chem* 280(21): 20740-20751.

Zhang, Y., J. Desharnais, *et al.* (2002). "Crystal structures of human GAR Tfase at low and high pH and with substrate beta-GAR." *Biochemistry* 41(48): 14206-14215.

Zhang, Y., M. Morar, *et al.* (2008). "Structural biology of the purine biosynthetic pathway." *Cell Mol Life Sci* 65(23): 3699-3724.

Zhao, C., Y. Kumada, *et al.* (2006). "Cloning, overexpression, purification, and characterization of *O*-acetylserine sulfhydrylase-B from *Escherichia coli*." *Prot Expr Purif* 47(2): 607-613.

Zhao, X. and W. A. van der Donk (2016). "Structural characterization and bioactivity analysis of the two-component lantibiotic Flv system from a ruminant bacterium." *Cell Chem Biol* 23(2): 246-256.

Zhou, M., E. Hernandez-Sanabria, *et al.* (2009). "Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies." *Appl Environ Microbiol* 75(20): 6524-6533.

Zwieb, C. and S. Bhuiyan (2010). "Archaea signal recognition particle shows the way." *Archaea* 2010: 485051.