The Influence of Horizontal Gene Transfer in the Metabolic Transformation and Origin of Higher Archaeal Clades

Inaugural-Dissertation

for the Attainment of the Title of Doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Düsseldorf

presented by

Alejandro Nabor Lozada Chávez
from Mexico City, Mexico

Düsseldorf, May 2018

From the Institute of Molecular Evolution
at the Heinrich Heine University Düsseldorf

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University Düsseldorf

Supervisor: Prof. Dr. William F. Martin
Co-supervisor: Prof. Dr. Martin Lercher

Date of the oral examination: 18 / 07 / 2018

Abstract

An evolutionary transition defines a major and remarkable process of change for a species. In prokaryotes, these transitions have been intensively promoted by the mechanism know as Horizontal Gene Transfer (HGT), which is defined as the interchange of genetic material between two different species. In the course of this thesis, HGT events between Archaea and Bacteria were quantified and characterized to understand, particularly, their evolutionary impact on three main subjects of interest: 1) on the origin and evolution of higher archaeal taxa, 2) on the evolution of the energy and methane metabolisms in Methanosarcinales, and its influence on the transitions between freshwater and saltwater environments, and c) on the origin and evolution of the operons coding for the aerobic respiratory complexes in Halobacteriales.

To that end, a total of 267,568 proteins from 134 archaeal genomes were analyzed to identify HGT events across 1847 bacterial genomes. Our findings show that interdomain gene transfers are highly asymmetric, that is, transfers from bacteria-to-archaea (imports) are more than fivefold more frequent than those from archaea-to-bacteria (exports). Furthermore, we found that the archaeal-specific gene families and imports that specifically occur in the archaeal clades, share the same phylogenetic histories across the 13 know archaeal clades. A comparative analysis of branches in both sets of trees (imports and archaeal-specific) showed that these bacterial acquisitions correspond to the origin of the 13 recognized archaeal taxa. Additionally, our findings show that these imports are involved in multiple metabolic functions that are implicated in key innovations during the major evolutionary transitions of the archaeal clades.

The findings of this thesis also demonstrate that Methanosarcinales genomes contain a large number of imports (10%–35%), whose potential bacterial donors are derived from clades such as Firmicutes, Clostridia, and Proteobacteria. The main roles of these imports are related to metabolic functions. Interestingly, my results shown that several HGT events have influenced the bacterial origin of the aceticlastic pathway in Methanosaeta and Methanosarcina species, and the expansion of the gene repertory involved in the electron transport chain in the family Methanosarcinaceae. Furthermore, some of these bacterial acquisitions (Rnf and Mrp complexes) have played, apparently, a fundamental role in the transition between freshwater and saltwater environments in Methanosarcinales. In agreement with this, we detect the complete absence in some Methanosarcinales of two hydrogenases (Frh and Ech) that are key to perform the hydrogenotrophic methanogenesis. Thus, I suggest that after the acquisition of Rnf and Mrp, the loss of Frh and Ech took place, originating the incapability of the same Methanosarcinales species to grow under hydrogenotrophic conditions and live in marine environments.

Finally, my findings also demonstrate that the five membrane-bound protein complexes representing the aerobic respiratory chain system are well conserved and organized as operons across Halobacteriales. However, a large number of paralogous counterparts were identified in all analyzed genomes. Interestingly, we observed an apparent random location of all operons representing the complexes across the Halobacteriales, which might reflect the continuous rearrangements reported in the genomes of this clade. My findings also confirm that the genes forming the complexes I-IV were acquired from bacteria, although some exports are observed in the complex V. These findings not only extend and confirm those of Nelson-Sathi et al. (2012), but also note that only 75% of the genes reported by their study are located within operon-like structures, while the remaining paralogous counterparts are widespread along the Halobacteriales genomes.

As we observed in our data, the evolution of Archaea has been shaped by a large number of bacterial acquisitions. The most inventive groups in this domain of life, Methanosarcinales and Halobacteriales, have taken advantage of the mutational source provided by HGT to explore metabolic and physiological transitions, and thus, to cope with new and extreme environments. Although most of the bacterial acquisitions are related to metabolic functions, the role of others also found in this study, e.g. informational or cellular functions, requires further research.

Zusammenfassung

Ein evolutionärer Übergang definiert einen bedeutenden und bemerkenswerten Veränderungsprozess einer Spezies. In Prokaryoten wurden diese Übergänge intensiv durch einen als Horizontalen Gentransfer (HGT) bekannten Mechanismus, der als Austausch genetischen Materials zwischen zwei verschiedenen Spezies definiert ist, gefördert. In der vorliegenden Arbeit wurden HGT-Ereignisse zwischen Archaea und Bakterien quantifiziert und charakterisiert, um ihre evolutionären Auswirkungen, insbesondere im Hinblick auf folgende drei Themen zu untersuchen: 1) auf den Ursprung und die Entwicklung höherer Archaetaxa, 2) auf die Evolution des Energie und Methan-Metabolismus in Methanosarcinales und seinen Einfluss auf die Übergänge zwischen Süßwasser- und Salzwasserumgebungen, und c) auf den Ursprung und die Entwicklung der Operone welche die aeroben Atmungskomplexe in Halobacteriales kodieren.

Zu diesem Zweck wurden insgesamt 267.568 Proteine aus 134 Archaea-Genomen analysiert, um HGT-Ereignisse über 1.847 bakterielle Genome zu identifizieren. Unsere Ergebnisse zeigen, dass Interdomänen-Gentransfers hochgradig asymmetrisch sind, das heißt, Transfers von Bakterien zu Archaeen (Importe) sind mehr als fünfmal häufiger als solche von Archaeen zu Bakterien (Exporte). Darüber hinaus haben wir herausgefunden, dass die archaealspezifischen Genfamilien und Importe, die spezifisch in den Archaea-Gruppen vorkommen, die gleiche phylogenetische Geschichte über die 13 bekannten Archaea-Gruppen teilen. Eine vergleichende Analyse der Äste in beiden Baumgruppen (Import- und Archaea-spezifisch) zeigte, dass diese bakteriellen Erwerbungen dem Ursprung der 13 anerkannten Archaea-Taxa entsprechen. Darüber hinaus zeigen unsere Ergebnisse, dass diese Importe in mehrere metabolische Funktionen involviert sind, die während der großen evolutionären Übergänge der Archaea-Clans mit Schlüsselinnovationen in Verbindung gebracht werden.

Die Ergebnisse dieser Arbeit zeigen auch, dass Methanosarcinales-Genome eine große Anzahl von Importen beinhalten (10%-35%), deren potenzielle bakterielle Donoren von Kladen wie Firmicutes, Clostridia und Proteobacteria abgeleitet sind. Die Hauptrollen dieser Importe beziehen sich eher auf metabolische Funktionen als auf andere operative oder informationelle Funktionen. Interessanterweise konnte ich auch zeigen, dass die verschiedenen HGT-Ereignisse tatsächlich die Entstehung und Entwicklung von Methan- und Energiemetabolismen beeinflusst haben. Während einige Exporte mit dem hydrogenotrophen Reaktionsweg in Verbindung stehen, zeigt eines der Hauptresultate, dass Importe bakteriellen Ursprungs den Acetylsäuresynthesesweg in Methanosaeta und Methanosarcina-Arten sowie die Erweiterung des Genrepertoires an Membrankomplexen, die in der Familie Methanosarcinaceae an der Elektronentransportkette beteiligt sind, unterstützen. Darüber hinaus spielten einige dieser bakteriellen Vererbungen (Rnf- und Mrp-Komplexe) offenbar eine fundamentale Rolle beim Übergang zwischen Süßwasser- und Salzwasserumgebungen in Methanosarcinales. In Übereinstimmung damit beobachten wir die vollständige Abwesenheit einiger Hydrogenasen (Frh und Ech), die für die Durchführung der hydrogenotrophen Methanogenese entscheidend sind, in einigen Methanosarcinales. Daher postuliere ich, dass nach dem Erwerb von Rnf und Mrp der Verlust von Frh und Ech stattfand, was, wie unter früheren experimentellen Studien berichtet, die Unfähigkeit dieser Methanosarcinales-Spezies hervorrief unter hydrogenotrophen Bedingungen zu wachsen. Insgesamt führten diese Gengewinn/-verlust-Ereignisse sehr wahrscheinlich dazu, die Elektronenquelle einiger Methanosarcinales in H2-abhängig (Ech) und H2-unabhängig (Rnf) aufzuspalten. Schließlich zeigen meine Ergebnisse auch, dass die fünf membrangebundenen Proteinkomplexe, die das aerobe Atmungskettensystem repräsentieren, gut konserviert und als Operons in Halobacteriales organisiert sind.

In allen analysierten Genomen wurde jedoch eine große Anzahl von paralogen Gegenstücken identifiziert. Interessanterweise beobachteten wir eine scheinbare zufällige Anordnung aller Operons die diese Komplexe in Halobacteriales repräsentieren, was die kontinuierlichen Neuordnung widerspiegeln könnte, die in den Genomen dieses Stammes berichtet wurden. Meine Ergebnisse bestätigen auch, dass die Gene, die die Komplexe I, II, III und IV bilden, aus Bakterien gewonnen wurden, obwohl im Komplex V einige Transfers von Archaebakterien zu Bakterien beobachtet

wurden. Dies erweitert und bestätigt nicht nur die Studie von Nelson-Sathi et al. (2012), sondern auch, dass nur 75% der Gene, die von der genannten Studie berichtet wurden, innerhalb von operonähnlichen Strukturen lokalisiert sind, während die verbleibenden paralogen Gegenstücke entlang der Halobakterien-Genome weit verbreitet sind. Wie unsere Daten zeigen, wurde die Evolution von Archaea durch eine große Anzahl von bakteriellen Akquisitionen geprägt. Die erfindungsreichsten Gruppen in diesem Lebensbereich, Methanosarcinales und Halobacteriales, haben die von HGT bereitgestellte Mutationsquelle genutzt, um metabolische und physiologische Übergänge zu erkunden und somit neue und extreme Umgebungen zu bewältigen. Obwohl die meisten bakteriellen Übernahmen mit metabolischen Funktionen in Zusammenhang stehen, bedarf es weiterer Forschung um Rollen wie z. B. Informations- oder Zellfunktionen, anderer, ebenfalls in dieser Studie entdeckter HGTs zu bestimmen.

Table of Contents

1

INTRODUCTION

Prokaryotes have inhabited our planet for at least 3.5 million years and their history is a story of continuous evolutionary transitions. These organisms harbor in their genomes the most extended reservoir for both old and modern metabolic pathways (Oren 2004). Also, they had continuously exploited ubiquitous genetic mechanisms to interchange their genetic information, acquiring thus, novel capabilities to contend in a myriad of different environments. One of such evolutionary mechanisms is Horizontal Gene Transfer (HGT), discovered more than 50 years ago (Akiba et al. 1960; Tatum and Lederberg 1947), which is defined commonly as the exchange of genetic material between two 'non-genealogical' species (Goldenfeld and Woese 2007). Although the relevance of HGT in eukaryotic evolution remains controversial (Boto 2010; Schönknecht et al. 2013), now it is widely accepted as one of the main evolutionary driving forces in prokaryotes (Boto 2010). Overall, HGT has provided of a better understanding on the evolution of all domains of life (see Figure 1.1).

In Prokaryotes, HGT has been extensively studied, mainly at the local phylogenetic scale. Thus, several estimations of horizontal gene transfer across bacteria have been reported. For instance, around the 81±15% of the prokaryotic genes, particularly in Bacteria, have been found to be involved in HGT (Dagan et al. 2008). In the last 10 years, an increase of the reports showing the role of HGT in the evolution of Archaea have provided a more comprehensive view of this domain of life (Garrett and Klenk 2007; Makarova and Koonin 2003). For example, HGT has been associated to the transformation of lifestyles and metabolic pathways, as well as to the origin of taxonomic groups (Boucher et al., 2003; Koonin et al., 2001), although the latest has provoked some controversial discussions (Becker et al. 2014). Two recent studies have revealed surprising obseravations on the origin and evolution of metabolic processes in two archaeal groups. On the one hand, the large phylogenetic scale analysis from Nelson-Sathi et al. (2012) report the bacterial acquisition of hundreds of genes by Halobacteriales in a single HGT event, which is suggested transformed the anaerobic and autotrophic ancient methanogen into the aerobic and heterotrophic modern group of Halobacteriales. On the other hand, a study carried out by Fournier and Gogarten (2008) shown that Methanosarcina species, laterraly acquired two bacterial genes that allowed the use of acetate to produce methane rather than CO_2/H_2 as substrate. The results suggest that these HGT events are ancient and make possible the expansion of an ancient methanogenic pathway (*i.e.* hydrogenotrophic) to another (*i.e.* aceticlastic) by using different set of enzymes. From both studies, two important obervations are particularly interesting, the large amount of shared genes between Archaea and Bacteria, and the significant influence that bacterial acquired genes have on the evolution of these archaeal groups by few or many genes acquired.

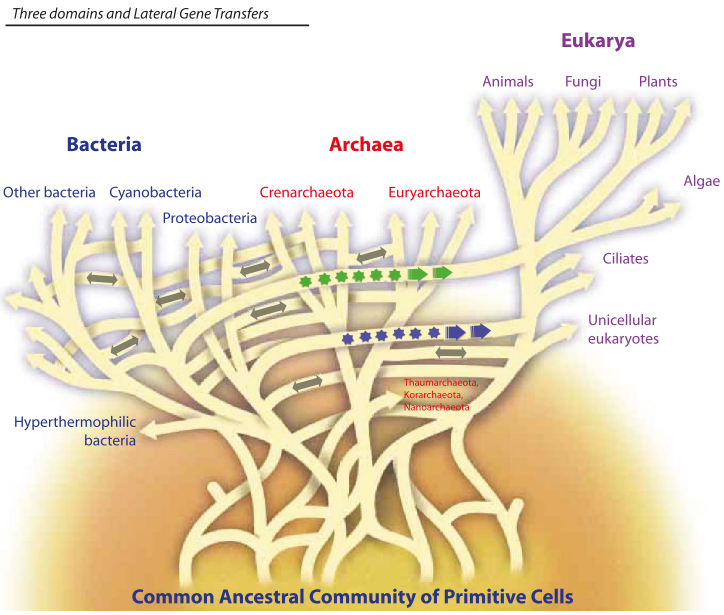


Figure 1.1. The tree of life and a schematic representation of HGTs in all species. Representation of species from different domains connected by an extensive network of links (HGTs) between branches non-related (e.g. black arrows, where the 'donor-recipient dynamic' differs by species or event). Despite of the widespread presence of HGTs, the species in the Tree of Life retains the tree-like structure (Doolittle 2000). Endosymbiotic events related to chloroplasts and mitochondria origins from bacteria to eukaryotes are shown with green and blue starts and arrows. Figure obtained and modified from Doolittle (2000).

1.1 Structure of the present work.

Based on these observations, therefore, the main goal of this research is to detect horizontal transferred genes between Archaea and Bacteria by the implementation of a large-scale phylogenetic analysis and comparative genomics. Methodologically, this study involves thre major goals: quantification and characterization of the horizontal transfers between Archaea and Bacteria; 2) the direction of the transfers, such as *bacteria-to-archaea* (imports) and *archaea-to-bacteria* (exports); and 3) the functional characterization of the HGT events across Archaea.

This analysis will be particularly focused in three major subjects: 1) on the origin and evolution of the traditionally known *archaeal clades*, 2) the expansions of the methanogenic pathways in *Methanosarcinales* and its influece on the aquatic transition between freshwater-saltwater environments, and 3) the origin and conservation of the membrane-bound protein complexes of the aerobic respiratory chain system in *Halobacteriales*. Thus, these three different studies were carried out during the present PhD research, and were divided in chapters as next:

1. Origins of major archaeal clades correspond to bacterial acquisitions (Chapter 6).
2. HGT events and gene losses have shaped methanogenic and aquatic transitions in Methanosarcinales (Chapter 7).
3. Protein complexes of the bacterial respiratory chain system in Halobacteriales are conserved and organized in an operon-like manner (Chapter 8).

In order to introduce the research of the present work, a short description of each Chapter is presented below.

1.1.1. Origins of major archaeal clades correspond to bacterial acquisitions (Chapter 6).

In this section, the origin of major archaeal clades was evaluated in the light of gene acquisitions through HGT from Bacteria. To assess this goal, an extensive phylogenetic analysis was performed through 8,779 archaeal protein families having bacterial homologs and 16,983 archaeal specific protein families (i.e. with non bacterial homologs). The direction of the HGT event was classified as *bacteria-to-archaea* (2,264 imports), *archaea-to-bacteria* (391 exports), and as *undefined* (4,074 non-monophyletic).

This information was employed as the basis for further analyses in this thesis (Chapters 7 and 8). First, our results show that there are several genes archaeal clade specific, and they were found in both set of trees, "imports" and "archaeal specific". The gene transfers between both prokaryotic domains is highly asymmetric, that is, transfers from bacteria-to-archaea are fivefold more frequent than from archaea-to-bacteria, suggesting that acquisition in archaea than bacteria, might be more likely in the nature. A statistical approach was developed to compare the phylogenetic trees of imports against archaeal-specific to know whether or not both sets of trees evolved along the same phylogeny from a single origin. Thus, from this comparative test, the results suggest that the origin of the 13 archaeal clades correspond to the acquisition of 2,264 genes from bacteria in each archaea higher clade, such bacterial acquisitions might have acquired either independently or through a single massive lateral transfers. Furthermore, genes identified at the origin of these archaeal clades might implicate the acquisition of metabolic functions from bacteria as key innovations. In any case, the findings of this work demonstrate for the first time the massive lateral gene transfer that has occurred in both prokaryotic domains.

Publication:

Nelson-Sathi S, Sousa FL, Roettger M, **Lozada-Chávez N**, Thiergart T, Janssen A, Bryant D, Landan G, Schönheit P, Siebers B, McInerney JO, Martin WF: **Origins of major archaeal clades correspond to gene acquisitions from bacteria.** *Nature* 517:77–80 (2015).

Contribution:

Bioinformatics analysis through multiple software packages and in-house PERL scripts: (1) *Collection of genomes and species filtering* for the construction of the database with 1,981 prokaryotic genomes (2). BLAST and reciprocal Best Blast Hits analysis, (3) and *clustering of archaeal protein families* using the Markov Cluster algorithm (MCL) to estimate candidate orthologs, (4) *filtering bacterial strains sequences* to reduce the proteins redundancy within each archaea protein family, (5) *reconstruction of phylogenetic trees* to detect HGTs, (6) *taxonomic organization* to the define archaeal and bacterial groups of all prokaryotic genomes analyzed, and (7) *annotation of gene functions* for to assign a functional role of the HGT genes. Finally, (8) the construction of a *long-arrange matrix* was built to compare homologs between Archaea and Bacteria; its graphical representation is shown in the first figure in the chapter 6.

1.1.2. HGT events and gene losses have shaped methanogenic and aquatic transitions in Methanosarcinales (Chapter 7).

Methanosarcinales are considered one of the most versatile groups among methanogens, and one of the archaeal groups with the highest proportion of bacterial sequenced counterparts. Our main question is to understand to what extent is HGT implicated in the expansion or acquisition of new

metabolic and physiological capabilities in Methanosarcinales. To tacke this question, I identified and characterized HGt events in 3128 protein families through extensive phylogenetic analyses, comparative genomics, and the use of orthologous gene annotations. Thus, Methanosarcinales genes involved in the (1) methanogenesis, (2) halotolerance, and (3) energy conservation were particularly analyzed. First, my results show that there is a differential gene acquisition of transfers across Methanosarcinales. For example, multiple bacterial acquisitions are genera specific (e.g. Methanosaeta or Methanosarcina), whereas others are shared among several Methanosarcinales genera. Furthermore, my findings support previous studies showing that Firmicutes and Proteobacteria are the main bacterial donors of these transfers. Also, genes involved with metabolic, cellular and environmental functions represent the most frequent acquisition in this methanogenic group.

From the 10 Methanosarcinales genomes initially analyzed (version June 2012), several HGTs were detected in genes related to energy conservation, halotolerance, and the three pathways of methanogenesis (hydrogenotrophic, aceticlastic and methylotrophic). Our approach also detect that the largest fraction of HGT events were identified as undefined transfers (57%), while a minor proportion represent imports (33%) and exports (10%). After a manual inspection of the phylogenetic trees, the direction of the several undefined transfers was resolved as a bacterial acquisition or as an export. In the hydrogenotrophic pathway, the manually inspected cases represent exports, whereas some HGTs were reclassified as imports in the aceticlastic and methylotrophic pathways. All together, key bacterial acquisitions might be involved in 3 evolutionary transitions of Methanosarcinales: 1) the transitions between freshwater and saltwater environments, 2) the expansion of the methanogenesis to use other substrates, and 3) the loss of the capability of growing under hydrogenotrophic conditions, such as in *M. acetivorans*, as probably result of the loss two hydrogenases playing a crucial role in this pathway.

Third, to test the robustness of the previous inferences in Methanosarcinales, a further *comparative genomics* analysis and a *survey in the literature* of physiological features of these species were performed for 24 new complete genomes in Methanosarcinales (version March 2017). On the one hand, orthologous genes involved in the methane and energy metabolisms in Methanosarcinales were identified, illustrating major changes (i.e. gain and losses genes) occurred at the membrane level. On the other hand, the experimental reports for 48 Methanosarcinales (species and strains) from the literature were analyzed to examinate the capability or incapability of growing under different methanogenic substrates. This analysis suggests that a large number of Methanosarcinales species cannot grow under hydrogenotrophic conditions. Hence, the impact of HGT is analyzed in this context. First, the loss in the ability of growing under hydrogenotrophic conditions in several Methanosarcinales species might be related to the loss of two key hydrogenases, the membrane energy-conserving hydrogenase (Ech) or the cytoplasmic F_{420}-reducing hydrogenase (Frh). Thus, the ability to grow under hydrogenotrophic conditions appears to be confined exclusively to some freshwater species of the Methanosarcina genus that retained both hydrogenases. Regarding to the aceticlastic pathway, my results show that the two enzymatic systems that active acetate in this pathway, the ACS and the ackA/pta systems, were the outcome of HGT events. In this chapter, I suggest that the origin of this pathway took place by the bacterial acquisition of the AMP-forming acetyl-CoA synthetase (ACS) to use acetate at the hydrogenoptrophic ancestor of Methanosarcinales. After this event, a secondary and exclusive bacterial acquisition of two enzymes with similar function that ACS, the acetate kinase A (ackA) and phosphate acetyltransferase (pta), occurred at the Methanosarcina genus. In some species, the low-efficiency of ACS to activate acetate was replaced by a more efficient set of enzymes with the same function (ackA/pta). Furthermore, the evidence

presented here suggests that the methylotrophic pathway is widely spread among saltwater and freshwater species. Despite the phylogenetic evidence corroborating the presence of HGT the bacterial origin of this pathway in Methanosarcinales via HGT remains unclear.

1.1.3. Protein complexes of the bacterial respiratory chain system in Halobacteriales are conserved and organized in an operon-like manner (Chapter 8).

As some reports claim, the aerobic respiration in Halobacteriales has been acquired by HGT (Boucher et al., 2003; Daubin et al., 2002; Kennedy et al., 2001; Nelson-Sathi et al., 2012). This system in Bacteria involves a large number of proteins organized in five membrane-bound complexes, which can be found clustered in operons. In this final section, my work is focused on investigating whether the conservation of this aerobic respiration system in 23 Halobacteriales genomes is preserved as clusters of genes organized as operons. In addition, the origin of each clustered operon-like structure for the five protein complexes was re-analyzed in the context of HGT. My findings were compared with those genes used in the study of Nelson-Sathi et al. (2012) to represent four out of the five membrane-bound protein comnplexes of the respiratory system in Halobacteriales.

To achieve this, 1,179 genes corresponding to 80 protein families previously reconstructed (chapter 6) were analyzed using the KO gene annotation system from the KEGG database. Next, the location for all genes involved in the aerobic respiration along the genomes was identified, and potential clusters of genes organized as operon-like were identified. First, for each of the five membrane-bound complexes, an operon-like organization was found on each of the 23 Halobacteriales genomes. All the complexes are well conserved in Halobacteriales, but the whole synteny of all complexes in the genome is poorly conserved. That is, they are, apparently, randomly located, which might be in agreement with the continuous rearrangements that the Halobacteriales genomes undergone during their evolution. It is important to note, however, that similar genomic locations of operons can be appreciated in strains of the same species.

The approach used in the entire research to detect the direction of the transfers classified many of these genes as non-monophyletic, and not as imports as it could be expected. However, a manual inspection of these gene trees corroborates the bacterial acquisition of most genes from the aerobic respiration chain system in Halobacteriales. Specifically, this bacterial acquisition occurred in four out of five of theses complexes (I, II, II and IV), whereas some genes of the complex V (A-type ATPase synthase) appear to be either exported to bacteria or free of HGTs. Among the five protein complexes, a large amount of gene copies were identified. Furthermore, the donors of these genes come from different bacterial groups, such as *Firmicutes*, *Proteobacteria*, *Actinobacteria*, *Nitrospirae*, and *Mollicutes*.

Finally, the comparative analysis between the conserved operons detected in this study against the candidate list of 213 genes reported by Nelson-Sathi et al. (2012), shows that only 159 genes were identified within operons. The remaining 54 genes were identified either as paralogs or were excluded of my analysis because the gene GI numbers were not longer present in the NCBI. Indeed, paralogs were detected in all membrane-bound complexes, but the high frequency occurred in the complexes I, II, and IV. Despite that the five membrane complexes are well conserved across Halobacteriales, several duplications or rearrangements events have occurred apparently. In conclusion, a crucial evolutionary transition from an anaerobic to an aerobic lifestyle from Bacteria in Halobacteriales was undoubtedly the outcome of horizontal gene transfers of the entire respiratory chain system.

2

HORIZONTAL GENE TRANSFER

2.1 The genetic mechanisms of HGT

The frequencies of HGT not only vary across gene families and species (Dagan et al. 2008; Koonin et al. 2001), but also the *mechanisms of horizontal gene transfer* are diverse (see Figure 2.1). Thus, there is not a specific trend for the presence of exclusive mechanisms between or within taxonomic groups. In prokaryotes, genes can be horizontally transferred among lineages through six or seven different mechanisms: *mobile elements*, *transformation*, *conjugation*, *transduction*, *gene transfer agents* (GTAs), and *cell fusion* (Mevarech and Werczberger 1985; Olendzenzki and Gogarten 2009; Popa and Dagan 2011). The other two known mechanisms, *intracellular endosymbiotic gene transfer* and *introgression*, are only associated with eukaryotic systems (Soucy et al. 2015). In **intracellular endosymbiotic gene transfer**, the genetic material of an endosymbiont (or later of an organelle) is incorporated into the host genome. It has been suggested that this type of transfer led to two major evolutionary transitions in the origin of eukaryotes, namely "origin of mitochondria and of chloroplasts" (Lane and Archibald 2008; Timmis et al. 2004). **Introgression** is represented by gene flow due to interspecies hybridization followed by backcrosses to one of the parents' species (Soucy et al. 2015). This mechanism has been observed in plants and humans, such as transgenic crops with grow proximity to non-domestical relatives (Stewart et al. 2003), and in introgressed gene, such as microcephalin from an archaic Homo lineage (Evans et al. 2006), respectively. Another mechanisms of transfer included **mobile genetic elements** (vehicles for HGT), such as plasmids, transposons, integrons, and integrative conjugative elements (ICEs), e.g. conjugative transposons (Olendzenzki and Gogarten 2009). **Transformation** involves the uptake of naked DNA from the environment, which is introduced in the cell through the membrane by a pilus system (Johnsborg et al. 2007). C**onjugation** is the transfer of DNA via plasmids, a process that is mediated by a cell-to-cell fusion or cell contact that forms a tunnel through which the DNA is transferred directly into the cell. **Transduction** is the acquisition of DNA after a phage infection. There are two types of transduction: generalized and specialized. The *generalized transduction* occurs when the cell is lysed after phage infection, and thus, any DNA segment of the host can be introduced in the phage by accidental packaging. These segments can be transferred to the next host infection by a very small minority of phage progeny. In the *specialized transduction*, the prophage fails to separate with precision from the bacterial chromosome due to imprecise excising of itself along with flanking host DNA segments. Thus stored viral and bacterial genes in the capsid can be inserted in the host genome during a future infection (Griffiths et al. 2000). *Gene transfer agents* are small and random genomic pieces of host stored in capsids, phage-like vehicles, produced by a nearby donor cell and released to the environment (Lang et al. 2012). These agents have been found in Archaea and Bacteria and have been considered as *gene delivery systems* able to be integrated into a host chromosome (Soucy et al. 2015). **Cell fusion**

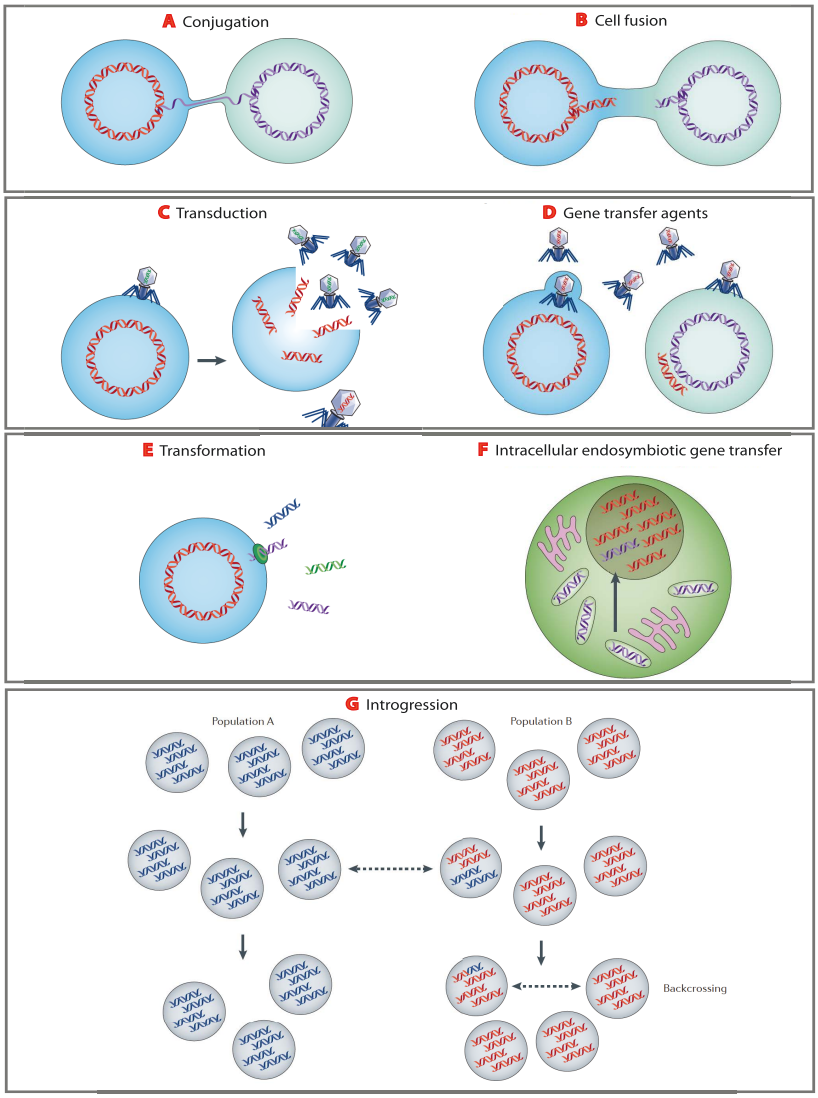


Figure 2.1. Mechanisms of gene transfer. Each panel represents different mechanisms of HGT. Seven HGT mechanisms are shown. **A**: Conjugation; **B**, Cell fusion; **C**, Transduction; **D**, Gene transfer agents (GTAs); **E**, Transformation; **F**, Intracellular endosymbiotic gene transfer; **G**, Introgression. *Conjugation* occurs through donor-recipient cell contact, while *Cell fusion* is a type of cell contact where transfers occur in a bi-directional form through a bridge formation between cells. *Transduction* is mediated by phages and can occur in either a generalized or specialized way (see text for more details). *GTAs* are driven by phages that no longer recognize their own DNA and only carry out fragments of host DNA. *Transformation* involves the uptake of DNA from the surrounding environment. *Intracellular* or *endosymbiotic gene transfer* occurs when the DNA from an endosymbiont (or organelle) is incorporated in the host genome. *Introgression* is described as a hybridization between two diverging species (orange and blue population in the figure). Backcrosses with one of the parent population (orange) can lead to only small piece of divergent genome (blue) remaining in the recipient. Figure from Soucy et al. (2015).

represents the formation of bridges between cells, formed by small aggregates through which a bidirectional exchange of DNA can occur. This process, described as a type of bacterial communication, has only been observed insome species of Archaea (Naor and Gophna 2013; Schleper et al. 1995) **Nanotubes**, on the other hand, produce a bridge between neighboring cells, which might o might not involved membrane fusion, and through which transfers of DNA and proteins can be conducted between two cells along with cytoplasmic content (Dubey and Ben-Yehuda 2011).

2.2 Natural barriers to HGT in prokaryotes

The ubiquitous frequency of HGT events observed in prokaryotes prompted the question of whether cellular systems have successful barriers against the influx of foreign DNA and genes. Such *natural barriers* can be placed in different stages of the genetic transfer and involve the roles of *the environment,* as well as of the diverse characteristics from the bacterial *donor* and *recipient*. For instance, Thomas and Nielsen (2005) have described a successful HGT process in four stages. The *first stage* involves the role of the **donors**, on which depending on the type of donor, the naked DNA can be either released and packaged into a phage particle or be released and integrated as a plasmid into a chromosome by interacting with a mating-pair formation apparatus. The *second stage* consists on the uptake of the DNA sequence, where binding sites can match with the naked DNA. The phage receptors or the *pilus* cases, the 'recipient specificity' is a crucial step in DNA integration. The *third stage* involves the role of the **recipient**, where the uptake and successful entry of foreign DNA in the cell will depend on how the restriction of cells and the anti-restriction systems of the donor will start to interact. The *final stage* is where the establishment is completed. In this step, the new gene is completely integrated, can be replicated, and is likely to be subject of other mutational processes (such as *homologous* or *illegitimate recombination)*, and thus, the new gene might persist longer.

Some studies in bacterial genomes suggest that dissimilar GC content between the donor and recipient genomes can be an important biological barrier for those genes subject to HGT (Thomas and Nielsen 2005). The beginning of the acquisition of genes, for instance, the GC content or codon usage bias in the recent acquisitions can represent a cost at gene expression level (i.e., the fitness on the functionality of genes). Accordingly, a recent research using phylogenomic networks shows that the difference in genomic GC content between donors and recipients is <5% for most (86%) of the connected donor-recipient gene pair (Popa et al. 2011).

Other relevant HGT barrier mechanisms include *recombination, DNA repair systems,* the *Restriction-Modification systems* (**RM system**), and the *Clustered Regularly Interspaced Short Palindromic Repeats-based system* (**CRISPR**). The **RM system** is widely present in bacteria and archaea, it comprises the enzymatic action of DNA methyltransferases (MTase) and restriction endonycleases activity (Rease). Rease recognizes and cleaves foreign DNA segments at specific sites, whereas MTase establishes the discrimination between *self* and *non-self* DNA by transferring methyl groups to the same specific DNA sequence within the host genome (Vascu and Nagaraja 2013). The **CRISPR system** consists of identical repeated sequences separated by highly variable DNA sequences (called 'spacers') derived from phage, plasmids (Mojica et al. 2005), *cas* genes, and small RNAs. These "*spacer*" sequences guide CRISPR-associated (cas) proteins to identify, recognize or silence foreign DNA containing complementary sequences (see Figure 2.2) (Makarova et al. 2015).

On the other hand, *homologous recombination* plays a role in the integration of the foreign DNA, which is constrained by DNA sequence divergence between the host and the foreign DNA. The functionality of the enzymes involved in homologous recombination affected by the DNA divergence (mainly in the initial steps) because such enzymes require a minimum length of 'similarity', also known as the minimum efficient process segment (MEPS) (Shen and Huang 1986). In spite of the large number of nucleotide differences or large insertions in the recombinant sequences, it is still possible for the recombination process to start (Bianchi and Radding 1983).

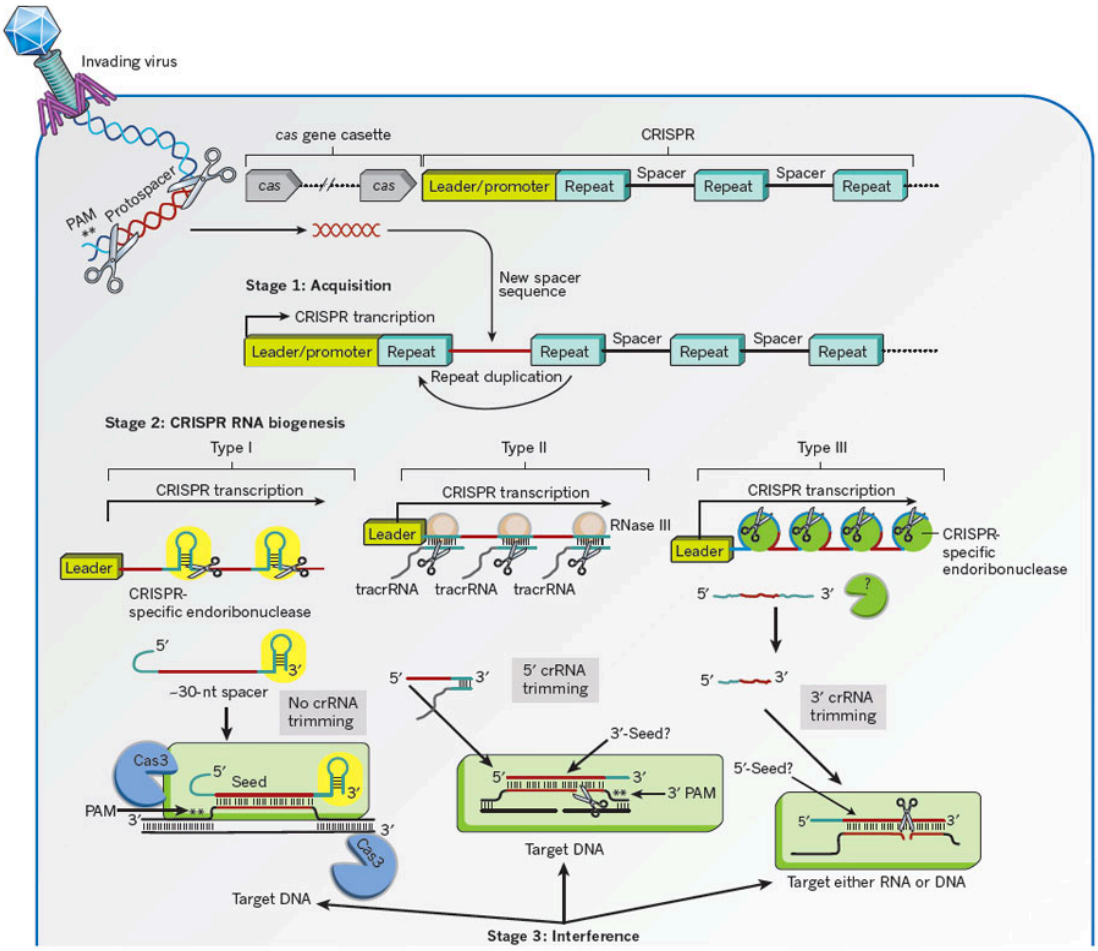


Figure 2.2 CRISPR-mediated adaptive immune system diversity in Archaea and Bacteria. The system is divided in three stages: Acquisition, Expression (or RNA biogenesis) and Interference. Stage 1: *Acquisition* represents the stage where multiples sets of CRISPR-associated (cas) genes (grey arrows) encode proteins required for a new spacer sequence acquisition. CRISPR locus consists of a series direct repeats separated by unique spacer sequences acquired from invading genetic elements (protospacers). Protospacers are flanked by a short motif called 'protospacer adjacent motif' (PAM,** upper left side). Stage 2: *Expression* stage: The biogenesis of CRISPR RNA is divided in three systems (I, II, III). In the system I and III, long CRISPR transcripts[1]) are processed into short CRISPR-RNAs (crRNAs), by the action of different endonucleases (yellow and green circles). In the type II system, a *trans*-acting antisense RNA (tracrRNA) with complementarity to the CRISPR RNA[2] repeat sequence forms an RNA duplex that is recognized and cleaved by cellular RNase III (brown ovals). On each system, the mature crRNA associates with one or more Cas proteins to form a surveillance complex (green rectangles). Stage 3: *Interference* stage: Type I systems encode a Cas3 nuclease (blue pie), which may be recruited by the surveillance complex after following target binding. A short high-affinity binding site called a seed-sequence has been identified in some type I systems and genetic experiments suggest that type II systems have also a seed sequence located at the 3′ end of the crRNA spacer sequence. Figure taken from Wiedenheft et al. (2012).

However, the methyl-directed mismatch repair system can detect the mismatches between nucleotides and stop the recombination if the sequence divergence is too high (Shen and Huang 1986; Worth et al. 1994).

[1] Complementary RNA strand from DNA.
[2] crRNAs have a length of 8 nucleotides upstream of each spacer sequence.

As commented previously, the environment plays an important role when the transfer occurs without the aid of mobile elements; for instance, when DNA is released to the environment and it is exposed to degradation by DNases. However, some plasmids are able to remain intact for long periods of time. In addition, the functionality of the DNA can be compromised by the actual physical characteristics of the ecosystem (Thomas and Nielsen 2005). On the other hand, *selectivity* in the type of DNA that recipients are able to translocate across the membrane has been described in species such as *N. gonorrhoeae* and *H. influenzae*, whereas non-selectively uptake of genomic segments in the species (Lorenz and Wackernagel 1994).

Once the genes acquired by HGT are successfully expressed in the new host's genome, they could represent a negative effect on the functionality of the cell, particularly those genes that might be highly connected in functional networks. Accordingly, there is a certain susceptibility on the type of genes that are successfully transferred by HGT, and they are divided in two big categories according to the *Complexity hypothesis*: 1) *informational genes* have several molecular interactions and usually involved in DNA replication, translation, transcription; 2) *operational genes* are those that have few interactions and are usually involved in cell maintenance and energy production. Based on this classification, the complexity hypothesis proposes that informational *highly connected genes* are not participating in HGT events with the same frequency as operational genes (Jain et al. 1999). Therefore, some reports have argued that the high connectivity of some proteins (given a protein-protein interactions) could denote another kind of resistance for HGT events (Cohan 2011; Price et al. 2008; Wellner et al. 2007). In *E. coli*, for instance, the integration of new genes in a regulatory network depends on the number of partners of the gene product, so that those horizontally acquired genes with low connectivity are localized in the periphery of the network and, in some cases, they can be easily erased from the recipient genome in the absence of a beneficial function (Lercher and Pál 2008).

2.3 Genomic detection of HGT events

Detection and interpretation of HGT events are crucial to understand how prokaryotic populations evolve through time, and comparative analyses have been employed to fulfill both tasks. In general, there are two types of methods to detect HGT: **compositional** or **parametric** and **phylogenetic** methods (Ravenhall et al. 2015; Zhaxybayeva 2009). The *compositional* methods search sections of a genome that are significantly different from the genomic average (anomalous) composition, which can be measured by different approaches. Such as **codon adaptation index** (CAI), **GC percentage** analysis, **Bayesian models**, and **higher order Markov models**. However, these approaches have been criticized because the difference in base composition and codon usage used between the possible transferred genes and the host sequences can be insignificant in some cases (Koski et al. 2001), leading to a compositional homogeneity among bacterial genomes. Nonetheless, these methods are frequently used to detect recent HGT events where compositional heterogeneity can be detected between donor and recipient (García-Vallve et al. 2000; Nakamura et al. 2004). Similarly, the *GC genomic composition* method can introduce a false signal in detecting HGT. This is because the signal can be **ameliorate**, that is, adopt the signatures of their new genome over the time (Lawrence 1997), reason by which this method is more reliable to detect recent HGT events. One of the most frequent analysis is the identification of HGT by high score similarity or identity, or the so-called '*significant homology*' determined by using the database search tool BLAST (Altschul et al. 1997). For example, identify a significant homology of a gene between distantly related species might be considered as a

rapid method to detect HGTs (Brown 2003). However, this score does not always reflect an evolutionary relationship between sequences (Brown 2003), meaning that the homology between the analyzed sequences cannot be assured.

The *phylogenetic methods* identify HGT according to the incongruent relationship detected in the phylogenetic trees between the candidate HGT gene and the species tree. To that end, this method relies on the reconstruction of trees for both individual gene families (i.e, a set of putative orthologous genes) and for a reference tree that attempts to represent the evolutionary history of the species under study through, for instance, genome trees, rRNA trees, concatenated genes tree (Zhaxybayeva 2009).

As with any phylogenetic reconstruction method owing to the phylogenies can produce potential pitfalls or spurious results of data quality, taxonomic sampling, analytical method and differential gene loss (Kurland et al. 2003). Also, the interpretation of evolutionary scenarios using phylogenetic analysis can be problematic due to the effect of secondary gene transfers, gene displacement and directionality of the transfer, as well as to the presence of independent gene transfer and differential loss. Furthermore, the loss and functional evolution of a gene can be related to the environment and lifestyle of the species rather than to HGT events, which makes difficult to determine the exact evolutionary of a gene (Huang and Gogarten 2006). Additionally, events such as lineages-specific losses, divergence, and unequal mutation rates in some taxa can produce similar evolutionary-relationship effects as those observed for HGT events. Despite these facts, phylogenetic trees based on protein families are still be the best method to detect either ancient or recent HGT events (a more detailed analysis will be discussed later on). That is because protein sequences change slowly in comparison to DNA sequences (Brown 2003).

2.4 HGT in prokaryotic genomes: a general overview

When compared to the role of vertical inheritance and other mutational mechanisms, HGT appears to explain a considerable fraction of the genome variation and complexity observed in Archaea and Bacteria (Darmon and Leach 2014; Jain et al. 2003; Koonin and Wolf 2008). Examples of other mutational mechanisms are gene loss, genome rearrangement, and gene duplication. Nevertheless, there is strong evidence suggesting that HGT might represent the most crucial component of prokaryotic genome evolution (Koonin and Galperin 2003; Koonin and Wolf 2008; Soucy et al. 2015). According to this, in bacteria and archaea the number of gene transfers has not only increased in the last year, but also a better understanding of the mechanisms of gene transfer has been obtained (Boto 2010; Koonin and Galperin 2003; Soucy et al. 2015). In both domains, for example, HGT can take place for a single or several genes, as well as *informational* and *operational* genes, although the frequency of HGT in the informational genes is minor, as the complexity hypothesis proposes (Jain et al. 1999). Nevertheless, cases of HGT in the ribosomal proteins have been reported in Bacteria (Brochier et al. 2000; Caro-Quintero and Konstatinidis 2015; Garcia-Vallvé et al. 2002) and Archaea (Yutin et al. 2012), as the product of ancient transfers, and other transfers between Bacteria and Archaea such as DNA polymerases (in the archaeon *Archeaoglobus fulgidus*), RNA phosphatase cyclase, DNA methylase, and rRNA gyrase.

Similarly, HGTs also involved the *transfer of complete operons* between Archaea and Bacteria (Hilario and Gogarten 1993; Koonin and Galperin 2003; Omelchenko et al. 2003; Price et al. 2006; Price et al. 2005). Examples of these are the nitrate reductase (narGHJI), urease (ureABCFG), formate hydrogen lyase (hyfBCDEFG), the Na^+-transporting NADH ubiquinone oxidoreductase (nqrABCEF),

the archaeal/vacuolar-type H^+-ATPase (ntpABCDEFGIK), and the Na^+/H^+ antiporter (mnhABCEFG). Despite these findings and the expected pressure to select operons, it seems that the rate by which complete operons are fixed after being horizontally transferred is lower in comparison to the origin of new operons via the transfer of individual or clustered genes (Lawrence 1997; Omelchenko et al. 2003).

HGT has also promoted metabolic expansions and innovations. For example, the production of B_{12} from glutamate (Woodson et al. 2003) and the production of methane using a different substrate (acetate) (Fournier and Gogarten 2008), represent two metabolic expansions. Another example of metabolic innovations is the acquisition of enzymes of the Methylaspartate cycle from different bacteria in Halobacteriales, which represent an alternative way to synthesize cellular building blocks (Khomyakova et al. 2011). As matter of this research, there are two lineages in archaea that are the main subject of study, the Methanosarcinales and Halobacteriales, in which cases of HGTs associated to this autotrophic methanogenic and heterotrophic halophilic groups will be commented later.

As shown in a recent study (Nelson-Sathi et al., 2012), prokaryotes share an impressive amount of horizontally transferred genes. In regard of this, previous reports have estimated the fraction of foreigner genes contained in bacterial genomes, which goes from small fractions (0.2%–7.2%) (Koonin 2009; Koonin et al. 2001) to higher ones (14%) (Dagan et al. 2008). However, there is a wide variation of this estimation among prokaryotes (Ochman et al. 2000; Ragan and Beiko 2009).

2.4 Impact of HGT in the bacterial genomes.

The frequent acquisition of genes via HGT in short periods of time has several repercussions on the evolution of prokaryotic genomes (Boto 2010; Popa and Dagan 2011). For instance, considerable changes in genome size can decrease the fitness of the recipient genome by decreasing gene expressions (San Millan et al. 2015). Also, the insertion of large non-coding regions or numerous genes can impose a cost in transcription and/or translation (Vogan and Higgs 2011). As a result of such consequences, HGT can redefine the ecological habitat and evolutionary route of the microorganisms (Cordero and Hogeweg 2009; Ochman et al. 2000). The functional role of such acquired genes dependent on each case and some genes are less likely to persist than others. So that, their maintenance in a genome is frequently associated with positive selection, to create, for instance, new functional alternatives (Ochman et al. 2000; Pal et al. 2005; Ragan and Beiko 2009).

Genes acquired horizontally can also increase and modify the gene repertory of the bacterial population. For instance, they can introduce novel traits, such as new metabolic pathways, the capability to synthetize or degrade compounds, *'antimicrobial resistance' or 'pathogenic islands*[3]*'*. The introduction of such novel traits promotes, in turn, the expansion of the niches available for an organism, as well as changes in a lifestyle. For instance, the transformation of a free-living organism to a pathogen, such as in the species *Listeria monocytogenes*, *Shigella flexneri*, *Bacteriodetes fragilis*, *Helicobacter pylori*) (Franco 2004; Gressmann et al. 2005; Nei 2003; Vazquez-Boland et al. 2001). On the other hand, HGT has a significant impact in estimating the core of dispensable genes (i.e., the pan-genomes size) that are essential and dispensable across species, which also makes difficult to

[3] Pathogenicity islands (PAIs) are distinct genetic elements on the chromosomes. PAIs encode various virulence factors and are normally absent from non-pathogenic strains of the same or closely related species (Gal-Mor and Finlay, 2006).

quantify the total amount of genes representing a specific species. Thus, the non-uniform distribution of the horizontally transferred genes observed in several studies indicates the flexibility of prokaryotic genomes throughout evolution (Medini et al. 2005).

3

AN OVERVIEW OF ARCHAEA

3.1 Phylogeny of an ancient prokaryotic domain

Archaea constitutes one domain of life with unique cellular, ecological, biochemical, evolutionary traits and pathways. Since their discovery, Archaea has been defined as prokaryotes, being quite similar to Bacteria in terms of cell structure. However, this view has changed based on the analysis of molecular data, mainly performed using the universal and small ribosomal RNA (rRNA) subunit (e.g. 16S) and protein-based trees (e.g. the universally conserved GTPases) (Woese et al. 1990). Thus phylogenetic reconstruction based on such molecular data has led to its classification as one of the three domains of life. Because the Tree of Life (ToL) is an unrooted tree, it is interesting to assess which of the three domains are more closely related, whether Archaea is truly a monophyletic group, and where the root of the Most Recent Common Ancestor (MRCA) is located (Fournier et al. 2011). Previously have suggested that Archaea is more closely related to Eukaryotes (Woese 2002). This assumption is based on the analysis of the location of the root of the tree, which should be placed either in the bacterial stem or within the Bacteria, implying that prokaryotes are a paraphyletic group[4] (Cox et al. 2008). There are two hypotheses, now on debate, explaining the relation between Archaea and Eukarya. The *three-domains hypothesis* (i.e., Archaea forms a monophyletic group), suggests a shared common ancestry between Archaea and Eukarya. On the other hand, the *archaeal-host hypothesis* (i.e., Archaea form a paraphyletic group) implies that Eukaryotes arose directly from an Archaea. Both hypotheses are illustrated in Figure 3.1. Another proposal known as the *Eocyte hypothesis* suggests the presence of a sister group between eukaryotes and the eocyte (or Crenarchaota) group. This hypothesis is particularly supported by extended genomic analysis that cluster several members of Archaea into the named TACK group (**T**haumarchaeota, **A**igarchaeota, **C**renarchaeota, **K**orarchaeota) and branch it with eukaryotes as sister group (Fournier et al. 2011; Williams et al. 2013) (see Figure 3.1).

Currently, a recent report has found a new archaeal candidate named *Lokiarchaeota*, and suggests (based on phylogenomic analysis) that eukaryotes are placed within Lokiarchaeota at the base of the TACK group, forming this monophyletic group. This study concludes that *Lokiarchaeota* is the closest relative to eukaryotes (Spang et al. 2015). Furthermore, it has been found that *Lokiarchaeota* contains more eukaryotic-like genes than other known archaea, suggesting that several main eukaryotic features could have derived from related archaeal ancestors (Embley and Williams 2015). Yet, further studies are needed to clear these phylogenetic relationships.

Since the fossil records show that eukaryotes appeared on Earth at least 1.8 million years ago, it is difficult to find any earlier phylogenetic signal for the shared genes between archaea and

[4] Group that contain some, but not all descendants of a common ancestor

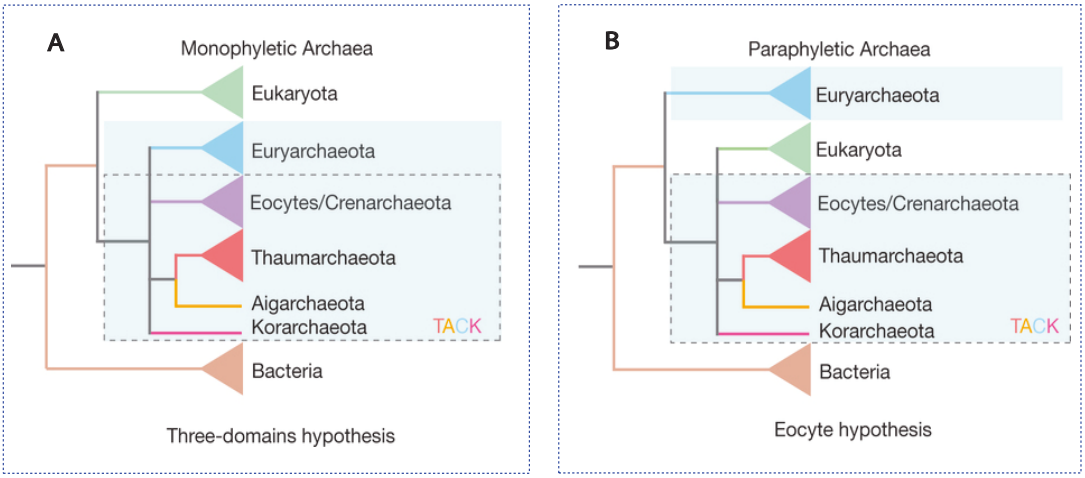


Figure 3.1. Hypothesis about the evolutionary relationships among the 'three domains of life' and 'Eocytes' hypothesis. *Panel A:* Schematic representation of the rooted *three-domains* tree of life: Bacteria, Archaea and Eukaryota. According to this hypothesis, Bacteria and Archaea are more closely related to each other because they share a common ancestor that is not shared with Bacteria. *Panel B:* Rooted *eocyte* tree that recovers the closest relative of Eukaryotes with the TACK super-group of Archaea (Taumarchaeota, Aigarchaeota, Crenarcarchaeota, Korarchaeota). TACK super group is defined in the squares in both panels (black dashes lines). Figure modified from Williams et al. (2013).

eukaryotes. Thus, the reconstruction of ancient events using phylogenetic trees presents several challenges to accurately assess accurate evolutionary relationships among the domains of life. For instance, some challenges are related to the **biological feature of the genomes**, such as the *function* and *differential evolution of genes* (Williams et al. 2013). Other challenges are related to some **methods of phylogenetic reconstruction**, including *parsimony* and *distance methods*, that do not take into account fast- and slowly-evolving sites nor the variation of the GC content from the sequences. Another artifact found in phylogenetic tree reconstruction is the *Long Branch Attraction* (LBA), where long branches are clustered together irrespectively of their evolutionary history (Felsenstein 1978); a problem more likely to appear when (mathematical) models ignore variations among DNA sites. Furthermore, *single-gene phylogenies* can show different conclusions when different criteria or method are used previous to the phylogenetic reconstruction. The contradictory conclusions of Katoh et al. (2001) and Ciccarelli et al. (2006) offer a good example. Both analyses reconstruct the phylogeny of the same set of 39 concatenated genes; however, each group used different alignment methods. Briefly, Katoh et al. (2001) used a specific order of the gene sequence in the alignment, which has been suggested to generate possible artifacts (Lake 1991), while Cicarreli et al. (2006) aligned the bacterial, archaeal and eukaryotic sequences separately before their concatenation in the alignment. This led to Katoh et al. (2001) to obtain a tree supporting the Eocyte hypothesis, whereas {Ciccareli, 2006 #75@@author-year} obtained a three-domain tree. In addition to this, single-gene trees, such as ribosomal and universal genes, are used sometimes to infer the evolution of the species. However, this approach can led to several artifacts and also hide the role of other mechanisms (e.g., HGT and endosymbiosis) in their evolution of species. After all, the genes involved in the reconstruction of species trees are only representing 1% of the genetic information of the species (Dagan and Martin 2006). Given the previous phylogeny reconstruction problems, new methods have been included in both maximum-likelihood and Bayesian frameworks, such as a relaxation in the assumption of homogeneous amino acid composition across sites and branches, which decreases the frequency of the LBA problem (Williams et al. 2013).

The origins of Archaea have been traced back to the Archaean era (>2500 m.a. ago), although other studies suggest much less time, ~1.6 m.a. ago. However, whereas the presence of two of the most ancient pathways suggest an earlier origin of Archaea: *methanogenesis* (>3.46 Gyr ago) (Ueno et al. 2006), and *sulfate reduction* (3.4 Gyr ago) (Shen et al. 2001). This data suggests that Archaea is very likely an ancient group and is divided in two major groups: Euryarchaeota and Crenarchaeota. Furthermore, it has been suggested that the ancestor of all extant Archaea was hyperthermophilic and that methanogenesis using CO_2 and H_2 is an ancestral trait of the Euryarchaeota group (Burggraff et al. 1991). It is interesting to note that the evolution of Crenarchaeota involves one particular moment of adaptation to mesophilic environments (mainly in aquatic habitats), while the evolution of Euryarchaeota involves the expansion of several groups owing to the adaptation to very different and diverse environments, including Thermoplasmatales, Halobacteriales and the methanogens Methanosarcinales (Chaban et al. 2005).

The origins of Archaea as unveiled through phylogenetic approaches are shaped by the rampant presence of HGT. However, only two well characterized set of genes are used to represent the Archaea domain, such as the *ribosomal genes* and *universal proteins*, and to reconstruct part of their evolutionary history. Assuming that both informational genes cannot or could not be transferred across all prokaryotes, according to the *complexity hypothesis*. The resulting tree represents a strong phylogenetic signal of ancient events, or at least it can be useful to retrace phylogenies of particular taxonomic groups or genes where horizontal gene transfer did not occur in principle (Gribaldo and Brochier-Armanet 2006). Recent reports using multiple genes (presumably part of the phylogenomic core in Archaea) and different models of tree reconstruction (e.g., maximum-likelihood or Bayes) have supported the division of the following major lineages within Archaea: *Eukryarchaeota* (including the groups *Methanopyrales, Methanobacteriales, Methanococcales, Thermoplasmatales, Archaeoglobales, Methanomicrobiales, Methanosarcinales* and *Halobacteriales*) (Garrity and Holt 2001b), *Crenarchaeota* (including the groups of *Thermoproteales, Desulfolobales and Sulfolobales*) (Garrity and Holt 2001a)*, Nanoarchaeota* (Huber et al. 2002), *Aenigmarchaeota*, *Diapherotrites*, *Nanohaloarchaeota* and *Parvarchaeota* (Rinke et al. 2013)*, Geoarchaeota* (Kozubal et al. 2013), *Bathyarchaeota* (Meng et al. 2014)*, Micrarchaeota* (Baker et al. 2010)*, Korarchaeota* (Barns et al. 1996) and *Thaumarchaeota* (Rosenberg et al. 2014). Other new archaeal groups have been reported, such as *Woesearchaeota* and *Pacearchaeota* (Castelle et al. 2015), *Lokiarchaeota* (Spang et al. 2015), and *Aigarchaetoa* (Nunoura et al. 2011)*,* which has been found living in terrestrial, marine, and subsurface thermal environments (Hedlund et al. 2015).

3.2 Metabolic diversity and habitats.

One of the major evolutionary differences between prokaryotes and eukaryotes is the capability of bacterial species to colonize extreme environments on Earth. Yet, Archaea holds genomic components similarly to eukaryotes that are not shared with the bacterial domain. For example, the presence of primases, helicases and polymerases in the DNA replication system (Olsen and Woese 1996). On the other hand, the size and organization of the genome, as well as the presence of polycistronic transcription units, and the use of specific sequences for the identification of the translation initiation site on the mRNA by the ribosome (Shine-Dalgarno sequences) makes Archaea more similar to Bacteria. Thus, the archaeal components shared with the other domains highlights the flexible nature of Archaea as a group that uses eukaryotic-like genes in a bacterial-like context (Gribaldo and Brochier-Armanet 2006).

Table 3.1. Biological, ecological and metabolic characteristics of each clade in Archaea, as compiled from the literature*.

Group[1]	Genera	Respiration	Metabolism	Habitats	Lifestyle
Korarchaeota	*Korarchaeum*	Anaerobic	Heterotrophic, probably obligate symbiont or 'scavanger'	Thermal habitats	Hyperthermophilic
Nanoarchaeota	*Nanoarchaeum*	Strict anaerobic	Obligate endosymbionts (of Ignicoccus hospitalis)	Sunmarine hot rocks, andy sediments, venting waters, material from black smokers, shallow marine areas, solfataric fields	Obligate symbiont
Thaumarchaeota	*Cenarchaeum* *Nitrosopopumilus* *Giganthauma* *Caldiarchaeum*	Aerobic	Mixotrophic, heterotrophic, chemoautotrophic. Putative symbiont of marine invertebrates	Marine waters, soils, hot springs, estuarine sediments, biofilms, caves	Ammonia oxidizers
Thermoproteales	*Thermoproteus* *Pyrubaculum* *Thermocladium* *Caldivirga* *Thermofilus*	Anaerobic (mostly), aerobic, and micro-aerotolerants	Facultative lithoautotrophic, and obligate heterotrophic	Springs, water and mud holes, soils of continental solfataric fields (low salinity and acidic to neutral pH)	Thermoacidophiles: Sulfur-reducer, energy production using organic and inorganic compounds
Desulfurococcales	*Desulfurococcus* *Aeropyrum* *Ignicoccus* *Staphylothermus* *Stetteria* *Sulfophobococcus* *Thermodiscus* *Thermosphera* *Pyrodictium* *Hyperthermus* *Pyrolobus* *Acidilobus*	Anaerobic (mostly) and aerobic	Mixotrophic, heterotrophic, lithoautotrophic, chemolithoautotrophic	Biotopes of volcanic activity land, shallow marine areas, hot springs, water or mud holes, soils of continental solfataric fields	Thermoacidophiles: Sulfur-reducing heterotrophs
Sulfolobales	*Sulfolobus* *Acidianus* *Methanollosphera* *Stygiolobus* *Sulfurisphaera* *Sulfurococcus*	Aerobic, facultative anaerobic	Mixotrophic, heterotrophic, Obligate lithoautotrophic, facultative chemolithoautotrophic	High-temp. continental solfataric areas, marine hydrothermal systems, terrestrial hot springs	Thermoacidophiles: Sulfate oxidizer, ferrous iron oxidizer, and sulfur-reducer (S^o autotrophy)
Thermococcales	*Thermococcus* *Pyrococcus* *Paleoccoccus*	Anaerobic	Heterotrophic, and obligate organotrophic	Terrestrial and submarine hot vents, offshore oil production platform, marine solfataras	Thermoacidophiles: Sulfur-reducing heterotrophs
Methanopyrales	*Methanopyrus*	Anaerobic	Autotrophic methanogen	Sediments in deep-sea hydrothermal vents	Hyperthermophile: Methanogens
Methanobacteriales	*Methanobacterium* *Methanobrevibacter* *Methanosphaera* *Methanothermobacter* *Methanothermus*	Anaerobic	Autotrophic methanogen	Freshwater and marine sediments, groundwater, rice paddy fields, anaerobic sewage digestors, others	Hydrogentotrophic methanogens. Some use formate, CO or secondary alcohols or H_2
Methanococcales	*Methanococcus* *Methanothermococcus* *Methanocaldococcus* *Methanotorris*	Anaerobic	Autotrophic methanogen	Marine environments, reservoir waters in oil fields, deep-sea hydothermal vents	Methanogens
Thermoplasmatales	*Thermoplasma* *Picrophilus* *Ferroplasma*	Facultative aerobic and aerobic	Autotrophic and heterotrophic	Hot springs, solfataric fields, continental volcanic areas, deep sea 'black smoker' vents, geothermal solfataric soils	Thermoacidophiles: Oxide iron
Archaeoglobales	*Archaeoglobus*	Facultative aerobic	Chemolithoautotrophic, Chemoorganotrophic	Terrestrial and submarine hydrothermal vents, hot oil fields water, hot springs	Thermoacidophiles: Sulfate reducers

Table 3.1. (Continued).

Clade	Genera	Respiration	Metabolism	Habitats	Lifestyle
Methanomicrobiales	*Methanomicrobium* *Methanogenium* *Methanolarcinia* *Methanoculleus* *Methanoplanus* *Methanofollis* *Methanocorpusculum* *Methanospirillum* *Methanocalculus*	Anaerobic	Autotrophic methanogen	Marine and freshwater environments, anaerobic digestors and the rumen, sewage sludge, industrial	Methanogens
Methanocellales	*Methanocella*	Anaerobic	Autotrophic methanogen	Rice fields soils	Methanogens
Methanosarcinales	*Methanosarcina* *Methanolobus* *Methanococcoides* *Methanohalobium* *Methanosalsum* *Methanohalophilus* *Methanosaeta*	Anaerobic	Autotrophic methanogen	Freshwater, marine environments, and extremely halophilic sediments, anaerobe sludge digestors and gastrointestinal tracts of animals, animal feces, industrial pipelines, fissure water of deep gold mines, and human periodontal pocket	Methanogens
Halobacteriales	*Halobacterium* *Haloarcula* *Halobaculum* *Halococcus* *Haloferax* *Halogeometricum* *Halohabdus* *Halorubrum* *Haloterrigena* *Natrialba* *Natrinema* *Natronobacterium* *Natronococcus* *Natronomonas* *Natronorobrum* *Halomicrobium* *Haobiforma* *Halosimplex* *Halalkalioccus* *Halovivax* *Halopiger* *Haloquadrata*	Aerobic and anaerobic	Autotrophic and heterotrophic	Wide variety of hypersaline environments: salt lakes, saltern crystallizer ponds, hypersaline soda lakes, soda lakes, acidic hypersaline environments	Halophiles

[1] Archaea has been recently divided in 15 phyla (Castelle et al. 2015), but only the most studied archaeal groups with their correspondent genera are shown.

* Biological, taxonomic and metabolic features were obtained from Schlegel and Jannasch (2006), Rosenberg et al. (2014), Reysenbach et al. (2001), Fendrihan et al. (2006), Angelidaki et al. (2011), Clingenpeel et al. (2013), Nealson (2008) and Steiglmeier et al. (2014).

The most distinguishing and unique feature in Archaea is represented by the membrane composition by phospholipids as isoprenoid ethers built on glycerol 1-phospathe (G1P) (Kates 1993). By contrast, the membrane composition of the other two domains is formed by fatty acid ester linked the stereoisomer 3-glycerol-phophate, G3P). And yet, the most impressive feature of Archaea is the diversity of metabolic pathways across the different lineages and groups. In particular, those metabolisms that have evolved in extreme environments, also known as *extremophiles*.

Despite the fact that several new archaeal clades remain uncultivated, several relevant biological traits have been detected in the current well-known lineages that harbor a great diversity of species and lineages, such as Crenarchaeota, Euryarchaeota, Nanoarchaea and Korarcheaota (see Table 3.1). For instance, **Nanoarchaeota** is a phylum of very small parasitic Archaea, where *Nanoarchaeum equitans* is a nano-sized, hyperthermophilic and strict symbiont that grows attached to the surface of an *Ignicoccus* species (hydrothermal archaea), forming thus an unique association (Huber and Kreuter 2014). **Korarchaeota** is a phylum of hyperthermophilic Archaea found in terrestrial hot springs

(Barns et al. 1994), in shallow marine hydrothermal vents (Marteinsson et al. 2001), hydrothermal vents fluids, chimneys (Takai et al. 2004), and on the sea floor of abyssal hills (Ehrhardt et al. 2007). In particular, an analysis of the *Candidatus Korarchaeum cryptofilum* genome suggests a physiology based on peptide fermentation, coupled with proton reduction to H_2 (Elkins et al. 2008). **Thaumarchaeota** is considered the most abundant archaeal lineage on Earth. Ecophysiological studies of ammonia-oxidizing Thaumarchaeota suggest adaptation to low ammonia and an autotrophic mixotrophic lifestyle (Pester et al. 2011). **Crenarchaeaota** is well-known for including a large number of hyperthermophilic and acidophilic species, which have been isolated from marine and terrestrial volcanic environments; additionally, their metabolic diversity involves a widespread range of chemoorganotrophic and chemolithoautotrophic organisms, from anaerobes to aerobes (strict or facultative) (Hershberger et al. 1996). **Euryarchaeota**, consists of strict anaerobic methanogens, extreme halophiles and hyperthermophilic species in Archaeoglobales and Thermococcales, while most acidophilic-thermophilic prokaryotes and cell wall free species belong to Thermoplasmatales (see Table 3.1). A large number of organisms in this kingdom produce methane as an integral part of their metabolism (Amils 2011), while members of the Halobacteriales group can live in environments with high concentration of salinity as heterotrophs (Oren 2014).

4

DIVERSITY IN METHANOGENS AND HALOARCHAEA

4.1 Methanogens and the different pathways for methanogenesis.

The production of methane is one of the most important biological processes on Earth, playing an important role in the global carbon cycle (Liu and Whitman 2008). Methane (CH_4) can be produced in the environment by *biological* or *geochemical systems*. On the one hand, the *abiotic synthesis* of methanogenesis is carried out in rocks by a set of **geochemical reactions** known as *serpentinization,* these reactions take place in a recently discovered vent system named *Lost City hydrothermal field* (LCHF) (Martin et al. 2008). In these systems, the serpentinization reactions produce H_2 that can reduce CO_2 to CH_4 geochemically. On the other hand, the *biotic synthesis* of methane is performed by a specific group of microorganisms called *methanogens,* which are only present in Archaea. The substrates used by methanogens to generate CH_4, on the other hand, can be obtained from different resources in the environment, for example, in **syntrophic a association,** that is, the combination of metabolic capabilities from two or more organisms to catabolize substances in a nutritional situation. These associations are between anaerobic bacteria, protozoa and/or anaerobic fungi, syntrophic bacteria, acetogenic bacteria, and methanogenic archaea (Thauer et al. 2008). A scheme representing such syntrophic associations is presented in Figure 4.1. Briefly, the biomolecules in these associations are hydrolyzed mainly into sugars, amino acids, purines, pyrimidines, fatty acids, and glycerol. Also, fermentative bacteria convert the present compounds in simple carbon acids (e.g. propionate, acetate and butyrate), alcohols (e.g. ethanol, propanol and butanol) and other compounds (ketones, CO_2 and H_2). The syntrophic bacteria use these products as substrates for growth, yielding acetate, CO_2 and H_2, as final products, which will be used by the methanogens (as substrates for growth) to convert them in methane (Thauer et al. 2008; Welte and Deppenmeier 2013).

Methanogens are strictly autotrophic anaerobes found in different anoxic environments, and they harbour a wide diversity of species (see Table 4.1). The main and unique metabolic ability of methanogens is to obtain energy from methane through the ancient metabolic *Acetyl-CoA pathway* (Ferry 1999; Stupperich et al. 1983). These microorganisms are exclusive to the Archaea domain, including phylogenetically close (e.g. within *Euryarchaea* phyla) or distant related clades. Figure 4.2 illustrates the major lifestyles and metabolic differences among all groups of methanogens. Overall, in Euryarchaeota, methanogens are divided in six recognized groups: *Methanococcales*, *Methanopyrales*, *Methanomicrobiales*, *Methanobacteriales*, *Methanocellales* and *Methanosarcinales* (Garcia et al. 2000; Liu and Whitman 2008; Rosenberg et al. 2014). In addition, recent findings have proposed a new class of methanogens based on the identification of new genomes with capabilities to produce methane, i.e. the *Methanomassiliicoccales,* which was previously referred as *Methanoplasmatales* (Borrel et al. 2013; Lang et al. 2014; Paul et al. 2012),

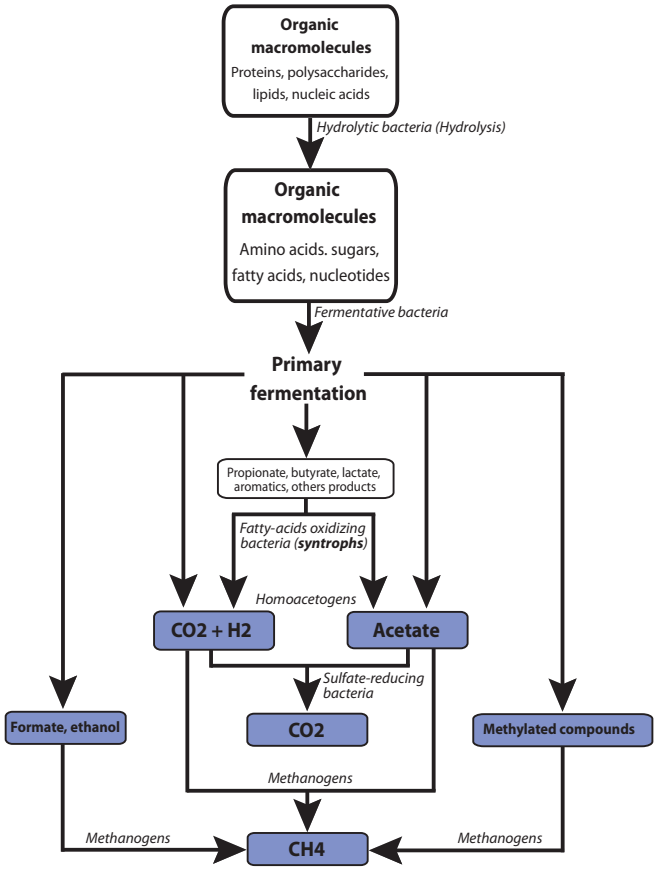


Figure 4.1. Overview of organic matter degradation to CO_2 and CH_4. The main complex organic compound is progressively broken down into simpler substrates by the cooperative interaction of a number of species growing in the ecosystem as part of a *syntrophic community*. The main substrates required to produce methane are present: CO_2 (hydrogenotrophic methanogenesis), acetate (aceticlastic methanogenesis) and methylated compounds such methylamines (mono, di, tri and tetra-methylamine), methane thiol and methylated sulfides (methylotrophic methanogenesis). H_2 and formate are the major electron donors to reduce CO_2 to CH_4, while ethanol is less used. Figure was modified from Hedderich and Whitman (2013).

and the *Methanonatronarchaeia* (Sorokin et al. 2017). Outside of Euryarchaeota, two new candidate groups have been suggested to play a role in the production of methane through specific enzymes, the *Bathyarchaeota* (Evans et al. 2015), and the *Verstraetearchaeota* (Vanwonterghem et al. 2016) (see Figure 4.2).

Currently, there are three main methanogenic pathways identified based on the type of substrate that are used by methanogens: *hydrogenotrophic, aceticlastic (*or *acetoclastic),* and *methylotrophic.* However, no all methanogens are capable to perform the three pathways (see Figure 4.2). There is conserved core of enzymes from the "hydrogenotrophic" methanogenesis in all six methanogenic lineages in Euryarchaeota (Bapteste et al. 2005). The names of such enzymes are: Formylmethanofuran dehydrogenase either in tungsten (Fwd) or molybdenum (Fmd) forms, Formylmethanofuran tetrahydromethanopterin formyltransferase (Ftr), Methenyltetrahydromethanopterin cyclohydrolase (Mch), Methylenetetrahydromethanopterin dehydrogenase (Mtd), Methylenetetrahydromethanopterin hydrogenase (Hmd), Methylenetetrahydromethanopterin reductase (Mer, coenzyme F_{420}), Tetrahydromethanopterin

Table 4.1. The seven-methanogenic archaeal clades and their general features from the literature.

Clade	Class*	Cell shape	Substrate (common)	Substrate (uncommon)	Motility
Methanobacteriales	H	Long-short rods, coccobacilli, spheres in paris, cylindrical rods, irregular-regular rods	H_2+ CO_2, H_2+ methanol	formate, alcohols	NR[2]
Methanopyrales	H	Long rigid rods	H_2 + CO_2	NR[2]	-
Methanococcales	H	Irregular cocci	H_2 + CO_2	formate	NR[2]
Methanomicrobiales	H	Rods, spirals, plates, irregular cocci	H_2 + CO_2,	formate, alcohols	- and +
Methanocellales	H	Rod, coccoid	H_2 + CO_2, Formate	NR[2]	NR[2]
Methanosarcinales[1]	H, A, M	Pseudosarcina, irregular cocci, sheathed rods	H_2 + CO_2, acetate, methyl compounds	formate is never used	+ - (freq)
Methanomassiliicoccales	M	NR[2]	Methanol	NR[2]	NR[2]
Bathyarchaeota	M	NR	Methanol, Methyl, Methylsulfide	NR	NR

* Pathways of methanogenesis.H: *hydrogenotrophic*; A: *aceticlastic*; M: *methylotrophic*.
Features and classification of the methanogenic groups were obtained from Thauer et al. (2008), Paul et al. (2012), Borrel et al. (2013), Evans et al. (2015), Rosenber et al. (2014). [1] Methanosarcinales is the only group of methanogens that possess cytochromes that play role in methanogenesis (Thauer et al. 2008). Similarly, this is the only group of methanogens comprising halophilic species, such as *Methanohalobium*, *Methanohalophilus*, and *Methanosalsum,* which live in hypersalines environments (moderately or extremely) (Oren 2011). [2] NR, *No reported*.

methyltransferase (Mtr), and Methyl coenzyme M reductase (Mcr) (Hedderich and Whitman 2006). The enzymes used in the "aceticlastic" and "methylotrophic" methanogenesis are different from those employed in the hydrogenotrophic pathway. The first one uses a kinase and phosphatase, whereas the second pathway uses a set of methyltransferases. Although the extant methanogenic pathways have different enzymes, they converge in the last step, consisting of conversion of methyl-coenzyme M (methyl-S-CoM) into methane (CH_4) by the enzymatic complex methyl-coenzyme M reductase (Mcr) (see Figure 4.3).

Briefly, the methanogenic pathways are described as follow. The **hydrogenotrophic methanogenesis** is the most ancient and widespread methanogenesis pathway in Euryarchaeota, and involves the reduction of CO_2 with H_2 (or formate) as an electron donor in seven steps. A schematic representation of this pathway is shown in the Figure 4.3 (see central section). In this pathway, methyl-S-CoM is produced through six reactions from CO_2 along with several intermediates (formyl-, methenyl-, methylene-, and methyl-coenzymes), and the electrons to reduce methyl-S-CoM to CH_4 derive from an external H_2 source.

The **aceticlastic methanogenesis** is performed by a limited number of Methanosarcinales members, mainly by organisms from the *Methanosarcina* and *Methanosaeta* genera that are capable of performing three enzymatic steps to cleaving the acetate (acetate activation), oxidizing the carboxyl-group to CO_2, and finally to reduce the methyl group to CH_4 (Welte and Deppenmeier 2013). *Methanosarcina* species prefer methanol and methylamine over acetate, and many species are able to use also H_2. *Methanosaeta* microorganisms are specialists, only capable of using

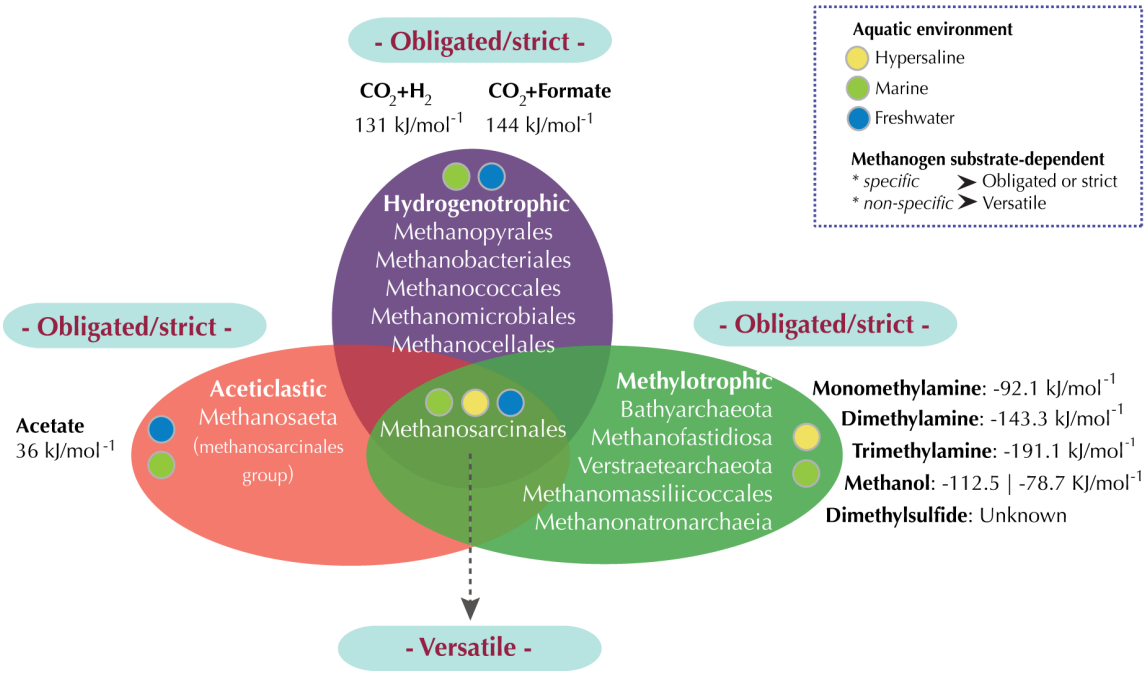


Figure 4.2. Aquatic environments, lifestyle, energy yielded by substrate and current classified methanogens in Archaea. Methanogens were grouped here based on three main features: type of methanogenesis, aquatic lifestyle, and their definition as *substrate specific* (obligated/strict) or *substrate non-specific* (versatile). Strict hydrogenotrophic (purple), aceticlastic (red), and methylotrophic (green) methanogens are shown within the ovals, while Methanosarcinales overlaps each methanogenesis pathway, indicating its versatility (at the center). Aquatic lifestyles, such as freshwater (blue), marine (green), and hypersaline (yellow), are shown in circles. Although methanogens are widespread in multiple environments, here an overview of their aquatic environments is shown (dashed blue frame). Environment information was obtained from Rosenberg et al. (2014), Paul et al. (2012), Borrel et al. (2013), Evans et al. (2015), Vanwonterghem et al. (2016), and Sorokin et al. (2017). Energy yielded by substrate is presented as energy free formation per molecule of substrate (kJ/mol^{-1}). Estimation of energy yielded by substrate, such as acetate, CO_2+H_2 or CO_2+Formate, and methylated C1 compounds, were obtained from Ferry (2002), Thauer (1998), Thauer et al. (2008), and Oren (1999), Oren (2002), respectively. Methanogens were grouped according to different traits among methanogens, such as niche, lifestyle, and energy yielded and the type of methanogenesis performed by each group. This information was obtained from Rosenberg et al. (2014).

acetate to perform methanogenesis, showing higher affinities for acetate (at low concentration) than *Methanosarcina* (requiring minimum concentrations about of 1mM) (Jetten et al. 1992). Accordingly, the first step of acetate activation is different in both Methanosarcinales genera (see left section at the dashed box in blue in Figure 4.3). On the one hand, *Methanosarcina* species activate acetate by ATP-dependent phosphorylation catalyzed by an acetate kinase (ack), and then a phosphostransacetylase (pta) converts the acetyl-phosphate in acetyl- CoA (named as *ack-pta* system) (see blue arrows). On the other hand, *Methanosaeta* species activate acetate using acetyl-CoA synthetase (ACS) (see orange dashed arrows), forming acetyl-CoA, and AMP pyrophosphate (PPi) (Deppenmeier et al. 2002; Jetten et al. 1992; Liu and Whitman 2008; Rosenberg et al. 2014; Welte and Deppenmeier 2013). After the acetate activation, the key reaction of this pathway is the cleavage of acetyl-CoA into its methyl and carbonyl moiety by the action of a CO dehydrogenase/acetyl-CoA synthase (CODH/ACS). Afterwards, CODH/ACS catalyze the reaction that releases CoA and transfers the methyl group first to H_4SPT, and then to CoM, where it will be reduced to CH_4 (Thauer 1998).

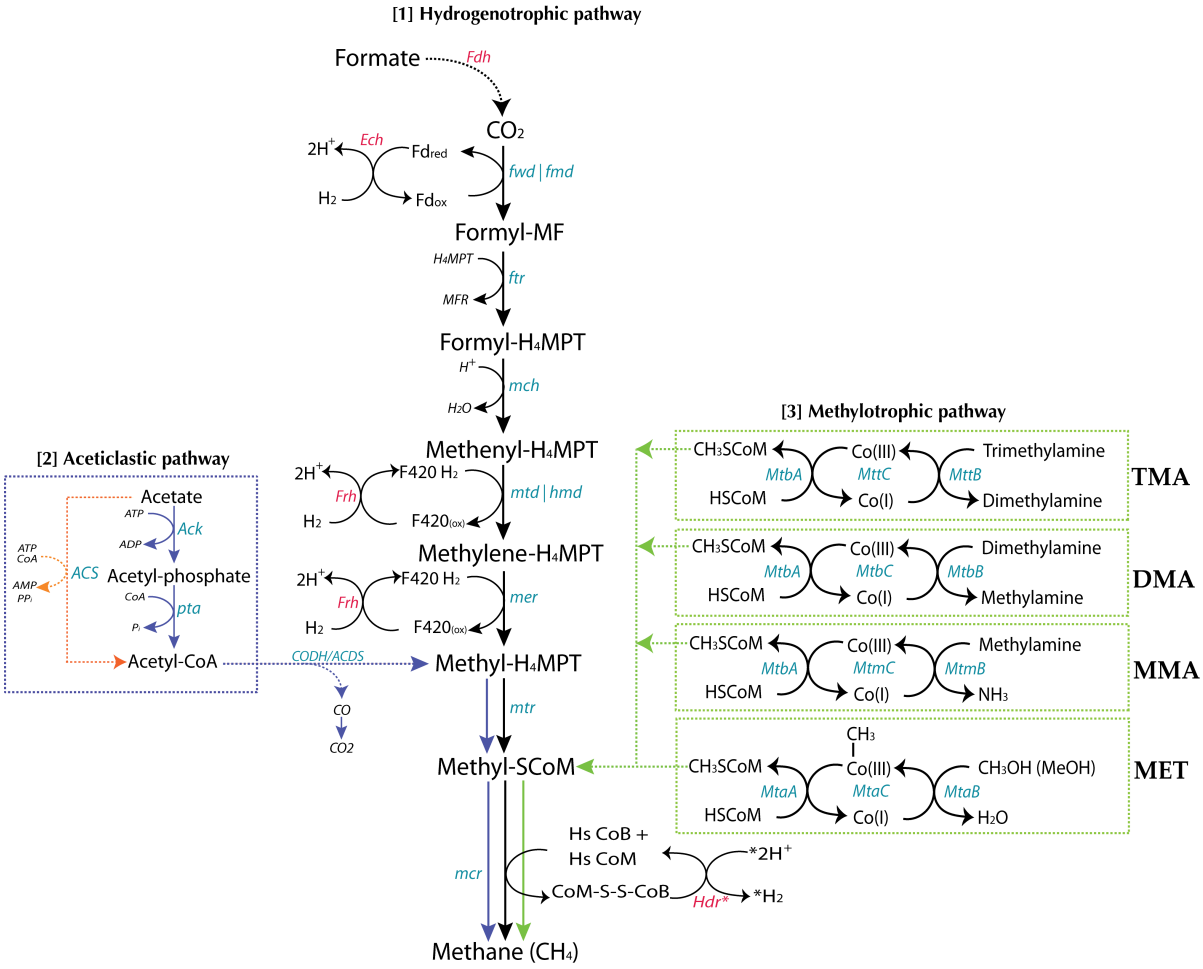


Figure 4.3. Methanogenesis overview: its three different biochemical pathways. The conversion of a methyl group is a common process in all methanogenic pathways; however, the source of methyl groups is varied. <u>*Central panel*</u> [1]: 'Hydrogenotrophic pathway' involves the reduction of CO_2 to methyl group as CH_4 through electron donors, H_2 or formate, 7 enzymes (colored in blue), and methyl-S-CoM. The gene names are as follow from top to the bottom (black downward arrows): Formylmethanofuran dehydrogenase, *fwd* (tungsten form) or *fmd* (molybdenum form); Formylmethanofuran-tetrathydromethanopterin N-formyltransferase, *ftr*; Methenyltetrahydromethanopterin cyclohydrolase, *mch*; Methenyltetrahydromethanopterin dehydrogenase, *hmd/mtd*; Coenzyme F_{420}-dependent N5,10-methylenetetrahydromethanopterin reductase, *mer*; Tetrahydromethanopterin S-methyltransferase, *mtr*; Methyl-coenzyme M reductase, *mcr*. <u>*Left panel (blue frame)*</u> [2]: 'Aceticlastic pathway' starts by activation of acetate to acetyl-CoA (blue box with dotted lines), yielding methyl and carbonyl groups, via two independent genera specific reactions in Methanosarcinales: *Methanosarcina* organisms use both acetate kinase (ackA) and phosphoacetyl transferate (pta) (blue arrows), while *Methanosaeta* organisms uses acetyl-CoA synthase (ACS) pathway (orange arrow). The methyl group is ultimately transferred to methyl-S-CoM, which is reduced to CH_4 (blue arrows in central panel). <u>*Right panel, (green frames)*</u> [3]: 'Methylotrophic pathway' involves the activation of methyl groups (obtained from different sources), by a methyltransferase and then transferred to CoM to be reduced to methane. This pathway use as substrates C-1 compounds containing carbon-bounded methyl groups to *oxygen*, *nitrogen* or *sulfur*. Depeding on the methyltransferase, these compounds included TMA, trimethylamines (*mttBC* and *mtbA*); DMA, dimethylamines (*mtbBC* and *mtbA*); MMA, monomethylamines (*mtmBC* and *mtbA*); MET, methanol (*mtaABC*). Other used compounds are tetramethylammonium (QMA), dimethylsulfide (DMS, *mtsABC*) or methane thiol. Membrane hydrogenases (shown in red): Formate dehydrogenase, *fdh*; Energy converting hydrogenase, *Ech* (*a* or *b*); Heterodisulfide reductase, hdr (* forms a complex with Mvh or Vho, *Rnf* or *Fpo*); and F_{420}-reducing hydrogenase *frh*. For simplicity, the production of reducing equivalents in this pathway is not shown (reverse reaction).

Finally, the **methylotrophic methanogenesis** is widespread in several species of the Methanosarcinales order, with the exception of Methanosaeta species (Lang et al. 2014; Paul et al. 2012; Rosenberg et al. 2014). As was shown in Figure 4.2, other new groups are also present, such as the *Methanomassiliicoccales, Methanonatronarchaeia, Verstraetearchaeota* and *Bathyarchaeota* (Borrel et al. 2013; Evans et al. 2015). An illustrative version of the methylotrophic methanogenesis is shown in Figure 4.3 (see green arrows and frames at the right section). In this pathway, diverse one-carbon (C_1) compounds are used, such as methanol, methylamines or methylated thiols. The methyl group of these substrates is first activated by the transfer to a corrinoid-binding protein in a reaction catalyzed by a specific methyltransferase, for example, the reaction MttB ⇒ MttC using trimethylamine (TMA) (see 'methylotrophic pathway' at left side in Figure 4.3). Then, the methyl group is transferred to CoM by a second methyltransferase, for example the reaction MttC ⇒ MttA (continued reaction using TMA in Figure 4.3), to be reduced to methane by methyl-coenzyme M reductase (mcr) (Thauer et al. 2008). Several methyltransferases are substrate specific and they a crucial role in this pathway. For instance, *methanol* (MeOH or MET) is substrate for mtaABC, *monomethylamine* (MMA) for mtmBC and mtbA, *dimethylamine* (DMA) is substrate for mtbBC and mtbA, and *trimethylamine* (TMA) is substrate for mttBC and mtbA.

4.2 Components and differences between methanogens in the energy metabolism: the membrane-bound electron transfer system.

Although the biochemical steps in the different types of methanogenesis are known, the interactions leading to the energy conservation processes (i.e. energy metabolism) are not well understood. As shortly described next, and illustrated in Figure 4.4, methanogens use membrane proteins complexes to transfer the electrons, although, this system is not the same in all methanogens. Based on these differences, methanogens are divided in two groups: methanogens with cytochromes (Methanosarcinales) and methanogens without cytochromes (the rest of methanogenic groups in archaea) (Thauer et al. 2008). Although this is known only for Euryarchaeota, future reports will confirm whether or not it also applies to all current methanogens in Archaea.

Some similarities can be observed among the hydrogenotrophic methanogens, for example, most of the enzymes involved in the reduction of CO_2 with H_2 are found in both methanogens with cytochromes and without cytochromes. However, the type of hydrogenases, and how they reduce the heterodisulfide (CoM-S-S-CoB) with H_2 represent the main differences between both types of methanogens (see Figure 4.4). For instance, the methanogens that depend on H_2 as an electron source are known as H_2-depedent or H_2-utilizing methanogens. By contrast, the methanogens that do not depend on H_2 as an electron source are known as H_2-independent methanogens. Also, the hydrogenases used for methanogens vary in number of subunits, as well as in their cellular location, which can be either at the membrane or in the cytoplasm. On the other hand, the heterodisulfide is reduced in methanogens with cytochromes by membrane complexes, while those methaongens without cytochromes use cytoplasmic complexes (Welte and Deppenmeier 2013).

Based on the above, there are four described membrane N_iF_e hydrogenases playing a role crucial role in the methanogenesis, they contain six core subunits (some contain extra subunits) and act in the reduction of ferredoxin with H_2. For example, the energy conserving hydrogenase *Ech* is present in methanogens that contain cytochromes, while the hydrogenases *Eha* and *Ehb*, and the membrane-

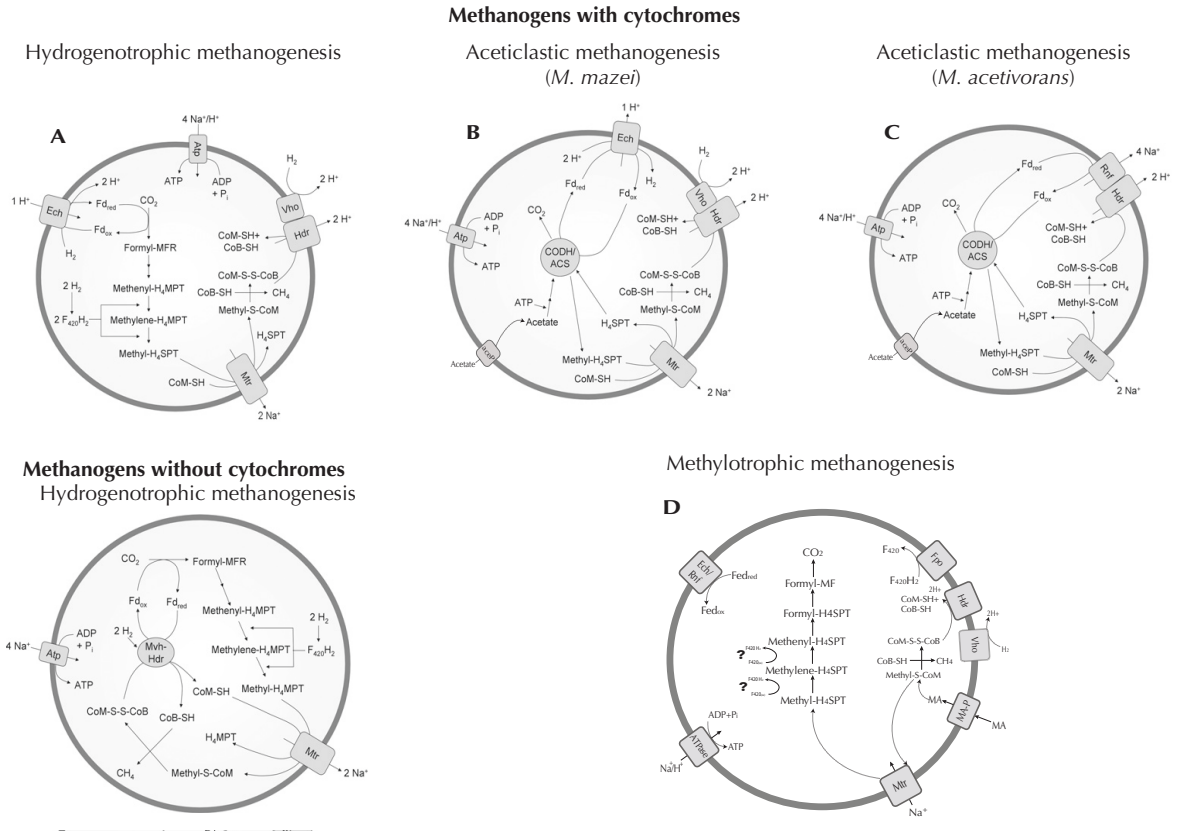


Figure 4.4. Proposed models of the methanogenesis in methanogens with and without cytochromes according to their different membrane-bound electron transfer systems. Methanogens are divided in two types based on the presence (A, B, C, and D) or absence of cytochromes (left side at bottom). Methanosarcinales represents the unique methanogenic group containing cytochromes in the electron membrane transport system, while the rest of methanogens do not, such as Methanopyrales, Methanomicrobiales, Methanobacteriales, Methanococcales, Methanocellales. This has been explored in methanogens within Euryarchaeota; however, evidence in other new methanogenic groups has not been shown. *Methanogens without cytochromes* (abbreviations): ATP or ATPase, A-type ATPase synthase; F_{420}, cofactor F_{420}; Fdox, oxidized ferredixin; MFR, methanofuran; Mvh-Hdr, complex of heterodisulfide reductase and the methyl Viologen-reducing hydrogenase; mtr, tetrahydromethanopterin methyl transferase. *Methanogens with cytochromes* (abbreviations): ATP or ATPase, A-type ATPase synthase; F_{420}, cofactor F_{420}; Fdox, oxidized ferredixin; MFR, methanofuran; Ech, energy-converting hydrogenase; Rnf, Rnf complex; Vho, (Methanosarcina) F_{420} non-reducing hydrogenase (dashed frame, absence or present in some species); Hdr, complex membrane Heterodisulfide reductase, Fpo, $F_{420}H_2$ dehydrogenase; Mtr, tetrahydromethanopterin methyl transferase, CODH/ACDS, carbon monoxide dehydrogenase/acetyl-CoA synthetase; aceP, acetate membrane permease transporter; MA-P, methylamines (permease) membrane transporter. *Versatility in Methanosarcinales*: the hydrogenotrophic methanogenesis pathway is mostly based on *Methanosarcina mazei*, while the aceticlastic methanogenesis pathway is based on two different Methanosarcina species, the hydrogenotrophic *M. mazei*, and the marine *M. acetivorans*. The schematic representation of the methanogenic pathways were adapted from Schlegel and Müller (2013), and the aceticlastic and methylotrophic pathways were modified based on Galagan et al. (2002), Welte et al. (2014), Rohlin and Gunsalus (2010). The model of the methylotrophic pathway is based on Welander and Metcalf (2005) and Spring et al. (2010).

bound hydrogenase *Mbh* are found only in methanogens without cytochromes. In addition, other hydrogenases are located in the cytoplasm, such as the hydrogenases (coenzyme) F_{420}-reducing hydrogenase (FrhABG)—which is present in both types of methanogens—and the methylviologen-

reducing hydrogenase (MvhADG), only found in methanogens without cytochromes. As mentioned earlier, the reduction of heterodisulfide is different in both types of methanogens. Methanogens with cytochromes use the membrane complexes the F_{420}-non-reducing hydrogenase viologen (*Vho* or *Vht*), and the heterodisulfide-reductase HdrDE complex. Instead, methanogens without cytochromes contain the cytoplasmic complexes MvhADG and HdrABC (a heterodisulfide-reductase). The *Vho* and *Vht* complexes are encoded by two operons, and both possess an additional subunit (vhoC or VhtC), which codes for a cytochrome b (Deppenmeier 1995). Similarly, has been found that a single methanogen of the Methanosarcinales order (*Methanosarcina acetivorans*) contained both HdrDE and HdrABC, which are encoded by two operons (Buan and Metcalf 2010), but only HdrDE contains a cytochrome b subunit (Heiden et al. 1994). Interestingly, this complex has been found exclusively in Methanosarcinales members (Thauer et al., 2008; Welte et al., 2013); however, the presence of both complexes, such as *M. acetivorans*, has not been explored in other Methanosarcinales species.

Methanosarcinales is the only methanogenic group containing cytochromes in the membrane electron transfer system, although cytochromes involved in other functions (no related to methane production) have been found in other archaea (Kletzin et al. 2015). Likewise, Methanosarcinales is the only group containing different membrane complexes from the rest of the other methanogenic groups. Considering this, it seems probable that major changes among methanogens have been exclusively taken place in Methanosarcinales, particularly at the energy and methane metabolisms (see Figure 4.4). Another trait that increase metabolic versatility of this group, is the presence of different membrane complexes involved in the electron transfer (Welte and Deppenmeier 2011; Welte and Deppenmeier 2013). Interestingly, has been suggested that this difference might be related to their life style (Spring et al. 2010) (see Figure 4.5). Accordingly, is well known that Methanosarcinales present three methanogenic pathways, through which ferredoxin is reduced either by a hydrogenase or by a Na^+ transporting NADH oxidoreductase. In other words, strict hydrogenotrophic methanogens use the energy conserving hydrogenase (i.e. Ech complex), while some methylotrophic methanogens, such as *Methanosarcina acetivorans* (Li et al. 2006) or *Methanohalophilus mahii* (Spring et al., 2010), possess a complex knows as Na^+ transporting NADH oxidoreductase Rhodobacter nitrogen fixation (i.e. Rnf complex). Interestingly, this membrane complex contains a cytochrome subunit and, apparently, was acquired from bacteria in *M. acetivorans* (Deppenmeier et al. 2002; Li et al. 2006). In contrast to this, *versatile substrate methanogens* (i.e. methanogens capable to growth in a wide range of different substrates), such as *M. mazei* or *M. barkeri*, use the Ech complex to performed methanogenesis (Welte and Deppenmeier 2013). Although this feature seems to be found only in species of the Methanosarcina genus (Welte and Deppenmeier 2011; Welte and Deppenmeier 2013), there is no report exploring this feature among other Methanosarcinales species. Moreover, and in contrast to the Methanosarcina species, Methanosaeta is a group of species known to be a *specific substrate methanogen* (i.e. growth specifically in acetate), and its energy metabolism, specifically the set of membrane complexes oxidizing ferredoxin, is unknown. These species lack of the complexes Ech and Rnf, instead, they have the $F_{420}H_2$ dehydrogenase complex (fpoABCDHIJKLMNO), but without the $F_{420}H_2$-oxidizing subunit FpoF, which makes these species unable to oxidize the cofactor $F_{420}H_2$ (Welte and Deppenmeier 2013). Evidence based on the analysis of *Methanosaeta thermophila* (Welte and Deppenmeier 2011), suggests that this strict aceticlastic methanogen channels the electrons of ferredoxin into a membrane-bound electron chain composed by a ferredoxin:methanophenazine oxidoreductase (Fpo-like complex) and the canonical methanophenazine-dependent heterodisulfide reductase (HdrED).

Table 4.2. Taxonomic classification and main biological features of the different genera in Halobacteriales, as obtained from the literature*.

Genera	Species name	Motility	Temp. (°C)	pH	NaCl %	NO_3 Reduc.	Quinones	Sample source
Halalkalioccus	*H. jeotgali*	-	21 - 50	6.5 - 9.5	10 - 30.0	-	NR	Shrimp jeotgai (Korean food)
Haloarcula	*H. californiae*	NR	NR	NR	NR	NR	NR	NR
	H. hispanica	+	25 - 50	NR	15 - 30.0	+	NR	Saltern, Spain
	H. marismortui	+	40 - 50	6.0 - 8.0	10 - 30.0	+	NR	The Dead Sea
	H. sinaiiensis	NR	NR	NR	NR	NR	NR	NR
	H. vallismortis	+	20 - 45	5.5 - 8.5	> 15	+	NR	Dearth Valley, Ca, USA
Halobacterium	*H. salinarum*	+	20 - 55	5.5 - 0.1	18 - 30	-	MK-8, 7	Salt lakes, salted food products
	H. sp.	+	20 - 55	5.5 - 0.1	18 - 30	-	MK-8,7, 6	Salt lakes, salted food products
Haloferax	*H. denitrificans*	-	30 - 55.0	6.0 - 8.0	9 - > 26	+	MK-8, MK-8(H2)	Saltern, California, USA
	H. mediterranei	+	> 45	6.5	> 30	+	NR	Saltern, Spain
	H. mucosum	-	30 - 55.0	6.0 - 10.0	10 - 30.0	-	NR	Hamelin Pool, W. Australia
	H. sulfurifontis	-	18 - 50	4.5 - 9.0	6 - > 30	+	NR	Saline sulfur spring, Okl, USA
	H. volcanii	+	30 - 40	NR	9.0 - 26	+	NR	The Dead Sea
Halogemetricum	*H. borinquense*	+	22 - 50.0	6.0 - 8.0	> 8	+	NR	Solar Saltern, Puerto Rico
Halomicrobium	*H. mukohataei*	+	40 - 50	6.2 - 8.0	15 - 26.0	+	NR	Salt flat, Argentina
Halopiger	*H. xanaduensis*	NR	28 - 45	6.0 - 11	15 - 30.0	-	NR	Salt Lake, Inner Mongolia
Haloquadrata	*H. walsbyi (C23)*	-	25 - 45	6.0 - 8.5	14 - 36.0	-	NR	Solar salterns, Spain, Australia
	H. walsbyi (DMS16790)	-	25 - 45	6.0 - 8.5	14 - 36.0	-	NR	NR
Halohabdus	*H. utahensis*	+	17 - 55	5.5 - 8.5	9.0 - 30	-	MK-8, type VIII-H2	Great Salt Lake, Utah
Halorubrum	*H. lacusprofundi*	+	4.0 - 45	NR	9.0 - 26	w	NR	Dee Lake, Antartica
Haloterrigena	*H. turkmenica*	-	45	NR	15 - 25.0	+	NR	Saline soil, Turkmenia
Natrialba	*N. magadii*	+	20 - 50	8.5 - 11	12 - 30.0	-	MK-8, MK-8(H2)	East-African alkaline lakes
Natronomonas	*N. pharaonis*	+	25 - 50	8.5 - 9.0	12 - 30.0	-	MK-8, MK-8(H2)	Alkaline lake, Inner Mongolia

* Biological information was compiled from multiple researches in (Oren 2006) and (Oren 2014).
The information shown in this table is related to Halobacteriales species only available in the year 2012 in the NCBI databank.

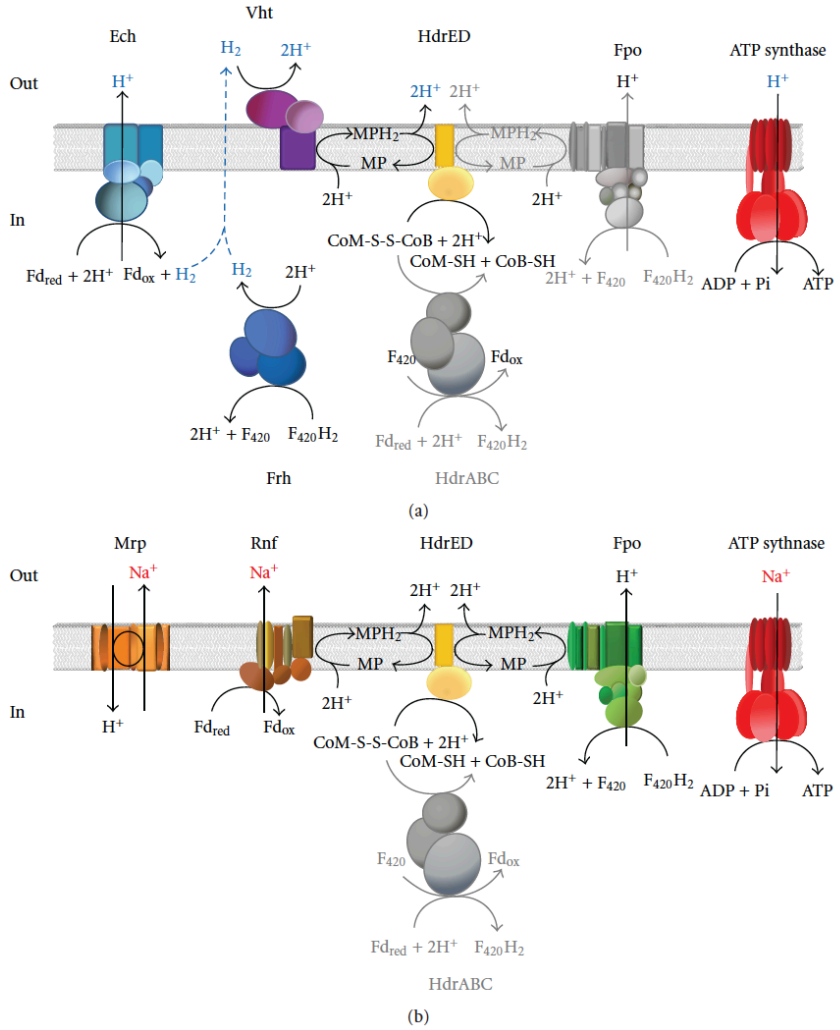


Figure 4.5. Freshwater and saltwater mechanisms of energy conserving electron-transfer in Methanosarcinales. Members of the family Methanosarcinaceae have a mechanism of energy conserving composed by different complexes, which can be found in (a) freshwater-inhabiting species, and in (b) saltwater-adapted species. The proposed of energy metabolism are mainly based on reports of gene expression and deletion in three *Methanosarcina* species: the freshwater *M. mazei* (Welte et al. 2010) and *M. barkeri* (Kulkarni et al. 2009), and the saltwater *M. acetivorans* (Guss et al. 2009; Li et al. 2006), and by comparative genomics in the hypersaline species *Methanohalophilus mahii* (Spring et al. 2010). Electron donors (e.g. hydrogen, $F_{420}H_2$, or reduced ferredoxin) depend of the substrate (CO_2, formate, acetate, methylated compounds). Figure was obtained from Spring et al. (2010).

4.3 Overview of the Halobacteria group

Extremely halophilic archaea, also known as Haloarchaea or Halobacteria (both terms are used indistinguishably in the literature), have developed strategies to couple with environments exposed to intense solar radiation, desiccation, high acidity, and high concentrations of salinity through sophisticated osmoregulation mechanisms (Oren 2002). Indeed, the minimal concentration of salinity that they require to grow is between 1.5 – 2.5 M NaCl (see Table 4.2). Haloarchaea belongs to the Euryarchaea phylum and presents different cellular shapes (e.g. circular, square and triangular), types of metabolisms, and also inhabit different ecosystems (Oren 2011). Notably, Haloarchaea possesses a number of eukaryotic-like genetic components, with no clear ancestral origins, such as RNA polymerases, transcription factors and TATA-box promoters.

Haloarchaea has other interesting traits:(a) a striking pigmentation (in a range of pink-to-red), due to the presence of carotenoids in their membranes and of the C50 compound bacterioruberin or its derivates, (b) a proton membrane pump bacteriorhodopsin for ATP production, which is only driven by sunlight, (c) non-polar properties of the lipids in the membrane, which provide protection against to sunlight, and (d) the formation of internal gas vesicles, which serves as floating devices to obtain oxygen (Oren 2014). Halobacteria are aerobic species, since they are able to use oxygen as terminal electron acceptor. However, some species can grow anaerobically when they are either exposed to light through nitrification or the use of energy obtained by the bacteriorhodopsin or during the dark through the fermentation of arginine (Hartmann et al. 1980).

It is interestingly to note that extreme halophilic heterotrops usually employ sugars as both carbon and energy resources, such as hexoses, pentoses, sucrose and lactose (Altekar and Rangaswamy 1992; Johnsen et al. 2001; Rawal et al. 1988). Extreme halophilic heterotrops that are not capable of sugar-degradation, like *Halobacterium salinarium* (Kauri et al. 1990) use amino acids and typical compounds of the salty environment instead, such as glycerol or organic acids excreted by some primary producers (Falb et al. 2008). Indeed, Haloarchaeabacteriales have different nutritional demands. While some species grow well on the presence of *simple compounds* (or single carbon sources) using ammonia as nitrogen source, other haloarchaea species require the presence of *complex compounds*, for instance, high concentration of yeast extract and other rich sources of nutrients have been observed in culture growth (Oren 2002).

4.4 Methanosarcinales and Halobacteriales: two groups with high bacterial genes-like in their genomes.

As described previously, Methanosarcinales and Halobacteriales represent groups with great diversity, physiologies and ecological traits. Some studies have related this diversity with HGT events, which apparently have shaped the evolution of both groups. For example, Nelson-Sathi et al., (2012) shows that both lineages contain a large number of genes acquired via HGT or, at least, are related to bacteria. Likewise, evidence has been presented for specific species or groups associated to these lineages (Allen et al. 2009; Deppenmeier 2002; Galagan et al. 2002; Garushyants et al. 2015). In Halobacteriales, for instance, has been suggested that several genes were acquired in a single massive event from bacteria, which transformed the anaerobic and autotrophic ancient methanogen of this group in an aerobe heterotrophic and halophilic organism (Nelson-Sathi et al. 2012). Despite the single event hypothesis, others studies suggest that such evolutionary transformations took place in multiple HGT events (Becker et al. 2014). In any case, there is no contradictory evidence of the large amount of bacterial-like genes contained in these Halobacteriales genomes. However, a limiting factor in previous studies is the number of genomes, used to perform the analyses. For instance, the study of Nelson-Sathi et al. (2012) only used 10 Halobacteriales genomes, while Methanosarcinales was represented by 4 species so far. This is probably because most of the studies are focused in Methanosarcina species (the most metabolic versatile methanogens in archaea) and at that time (NCBI database version of June 2012 contains 10 genomes only).

Given the increasing number of complete sequenced genomes in prokaryotes, a systematic and large-scale analysis of horizontal gene transfer in Methanosarcinales and Halobacteriales opens the possibility to understand and clarify the impact of HGT in the metabolic flexibility and evolutionary patterns belonging of these groups. Therefore, some of the main questions pursued in this thesis are:

On the origin of archaeal clades through the gene acquisition from bacteria:

- *How many genes in archaea have been acquired through HGT from bacteria?*
- *How many genes have been transferred to bacteria from archaea?*
- *What are the specific functions of the genes acquired horizontally across and on each archaeal taxa?*
- *Can the lineage-specific genes, derived from vertical and lateral transfer, describe the same phylogenetic origin of each archaea lineage?*

On the relation between HGT and the metabolic versatility in Methanosarcinales:

- *How many bacterial genes were acquired in Methanosarcinales?*
- *What are the most relevant metabolic pathways involved in these HGT events?*
- *Are the two recent methanogenic pathways and the versatile set of membrane-bound complexes the product of HGT events from bacteria?*
- *Are the core genes of the ancient hydrogenotrophic pathway involved in HGT events?*
- *Can certain genomic changes (such as gene gain/loss and HGT) explain some key ecological expasions (i.e. freshwater and saltwater) in Methanosarcinales?*

On the bacterial acquisition of the aerobic respiratory chain and its genomic organization in Halobacteriales genomes:

- *Are the genes of each membrane-bound protein complex organized as operon in Halobacteriales?*
- *Is the organization of the whole respiratory chain system in the genome conserved?*
- *To what extend is the respiratory chain system the outcome of HGT events in Halobacteriales?*
- *Are the HGT events potentially detected in Halobacteriales the outcome of a single or multiple evolutionary events?*

OBJECTIVES

The central objective of this work is the detection and analysis of HGT events with Bacteria to Archaea, with a particular focus on those acquired genes in Methanosarcinales and Halobacteriales particularly focused on those genes acquired in Archaea, as well as their functional characterization. Accordingly, to that end, three specific objectives are described next:

On the origin of higher taxa in Archaea:
Identify all HGT events and classify the direction of the transfers as bacteria-to-archaea (*i.e.* imports), archaea-to-bacteria (*i.e.* exports), as well as group all those archaeal-specific genes free of HGTs (*i.e.* archaeal-specific).

For each clade in archaea, perform a comparative analysis of two different sets of phylogentic trees (clade-specific and imports) to test whether both sets of trees evolved along the same phylogeny, corresponding to the same origin iof the archaeal group.

On the origin and evolution of the metabolic versatility in Methanosarcinales:
Characterizing the gene function and donors of all HGTs, either as gene transfers bacteria-to-archaea (imports) or to archaea-to-bacteria (exports). The identification of genes involved in the *methanogenic versatility* and *energy metabolism* of this group based on reference genes (acconrding to the literature) and evaluate: 1) whether their origin is related or not to HGTs events from bacteria, and 2) the biological contribution of these gene transfers along the complete sequenced Methanosarcinales species.

On the bacterial acquisition and genomic organization of the respiratory chain system in Halobacteriales:
Identifying all genes involved with the *five* membrane-bound protein complexes of the respiratory chain system in Halobacteriales, and evaluate: 1) whether each set of genes belonging to each complex is organized in an operon-like, and 2) whether or not the origin of these five complexes are the outcome of HGTs events from bacteria.

5

METHODS

Collection of genome dataset, gene annotation and species taxonomy. We analyzed the evolutionary impact of HGT from bacteria in 134 complete archaeal genomes. Within this sample, additional analyses were performed in two clades, Methanosarcinales and Halobacteriales. To that end, we downloaded a database of 1,981 complete sequenced prokaryotic genomes: 1,875 genomes were downloaded from the National Center for Biotechnology Information (NCBI, version of July 2012, ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/) (Altschul et al. 1997), and 6 recently sequenced haloarchaeal genomes (Lynch et al. 2012) were taken from the Joint Genome Institute (JGI) databases (Grigoriev et al. 2012) (https://genome.jgi.doe.gov). In total, 134 genomes belong to Archaea and the remaining 1,847 genomes belong to Bacteria. The taxonomy classification for all genomes was obtained from the NCBI taxonomy website (https://www.ncbi.nlm.nih.gov/taxonomy). The complete gene annotation for each genome was obtained from the NCBI genome files previously downloaded via RefSeq database. An additional filtering of genomes was performed for some analyses as is describing in the following sections.

Estimating protein families and homologs candidates across archaeal genomes. To identify putative homologous proteins across the 134 archaeal genomes, an archaeal all-against-all BLAST was carried out for all the 309,128 proteins. We consider candidate homologous proteins by using the reciprocal best BLAST hit (rBBH) approach (Tatusov et al. 1997) through an in-house *PERL script* and the following thresholds: local identity ≥25% and an e-value $\leq 1x10^{-10}$. A total of 11,372,438 pairs of rBBHs were obtained from this approach. A further improvement in the quality of the sequence alignment was performed by using a pairwise global identity (≥25%), through the Needleman-Wunsch algorithm available from the EMBOSS package (version 6.6.0.0) (Rice et al. 2000). By using the Markov cluster (MCL) algorithm (with default parameters), we obtained 25,762 clusters or families of archaeal proteins with ≥ 2 members with a *global identity* ≥25%, as "*edge* weights" in the clustering procedure. In order to include bacterial counterparts into the archaeal protein families, were detected candidate bacterial homologs to any member of the archaeal protein family by using the rBBH approach, as described above, and the following parameters: global identity ≥25% and an e-value $\leq 1x10^{-10}$. Because multiple bacterial genomes are represented by a high number of strains (see Supplementary Figure SF1), a *redundancy analysis* was performed on each archaeal cluster to include only the most similar bacterial hit from the available strains of the same species; the remaining bacterial sequences were removed. It is important to note, however, that several close member copies in a single genome—either archaeal or bacterial—, *e.g.* in-paralogs, might be present in a particular protein family.

Reconstruction of phylogenetic trees for archaeal protein families. Protein families with at least four members were aligned using the MAFFT program (version v6.864b) (Katoh and Frith 2012). Maximum likelihood trees for each family were reconstructed using RAxML (version 7.4.0) with the

PROTGAMMA+I+WAG model (Stamatakis 2006). In order to reconstruct an evolutionary and systematic organization of archaeal lineages, a *reference tree* was also reconstructed from a weighted concatenation alignment (Guindon and Gascuel 2003) of 70 universal genes present in single copy in all archaeal genomes used in this study.

Detection of horizontal gene transfer events. To detect possible cases of HGT events between Archaea and Bacteria, we first identified all protein families where 2 or more sequences of Archaea were present and, subsequently, all the phylogenetic trees were reconstructed. Second, on each protein family *archaeal clades* were identified based on the reconstruction of a "reference archaeal tree" (described previously) and their current taxonomy, whereas *bacterial clades* were identified according to their taxonomic classification based on the 'tree of life' reconstructed by (Ciccarelli et al. 2006). The groups of bacteria that could not be classified with the previous approach (because they are novel, recently updated or with no annotation in our dataset) were classified in a single predefined group designed as "others". Third, all phylogenetic trees with HGTs were classified according to the branching topology of the archaeal and bacterial clades (i.e. bipartition in the tree) through the use of an in-house PERL script and the *Newick utilities* package (Junier and Zdobnov 2010). This classification is described based on two large groups of trees: *monophyletic* and *non- monophyletic*, and they are described as follow. We coined the term "monophyletic" to refer to the HGT event particularly located within a single archaeal clade (i.e. in a single branch of the tree) in a protein family having bacterial homologs. That is, an archaeal clade is monophyletic if there exists a bipartition (branch) in the tree that separates all archaeal from all bacterial leaves (see *monophyletic* tree topology at left side in Figure 5.1). On the other hand, the term "non-monophyletic" was used to describe those cases in which the protein family is formed by ≥2 archaeal clades (see *non-monophyletic* tree topology at right side in Figure 5.1).

For the analysis of the HGTs in Methanosarcinales, the term "monophyletic" was used to describe two topologies in the tree defining HGTs. The first topology involves a bipartition in a tree that separates all Methanosarcinales from all bacterial leaves, and recalls the absence of other archaea. The second topology involves a bipartition in a tree separating bacterial leaves from both Methanosarcinales and other archaea leaves, i.e. a tree containing ≥2 archaeal clades (see *monophyletic* tree topology at left side in Figure 5.1). Similar to the analysis in all archaeal clades, here the term "non-monophyletic" was used to describe all trees where ≥1 archaeal clades do not form a monophyletic group.

Defining the direction of the HGT events between Archaea and Bacteria. The direction of a HGT event between the sequences of two domains (archaea and bacteria) was estimated by following the 'out of Africa' principle, which suggests that the taxon with the most diverse representation (or the most crowded taxon) of a given family sequence is the most likely source of origin (Cavalli-Sforza 1998; Koonin et al. 2001). Thus, the most crowded taxon (i.e. with the highest density) in a protein family was identified with the maximum number of clades representing either Bacteria or Archaea. To this purpose, the assignment of prokaryotic clades was performed with two methods. *Archaeal clades* were identified through a taxonomic classification and by the reconstruction of the "reference archaeal tree", as described previously, whereas *bacterial clades* were identified according to their taxonomic classification based on the recently updated 'tree of life' (Ciccarelli et al. 2006). The bacterial groups that could not be classified with the previous approach (because they are novel, recently updated or non-annotated groups in our dataset) were classified in a single predefined group as "others". Second, for each protein family the following measures were calculated to determine the

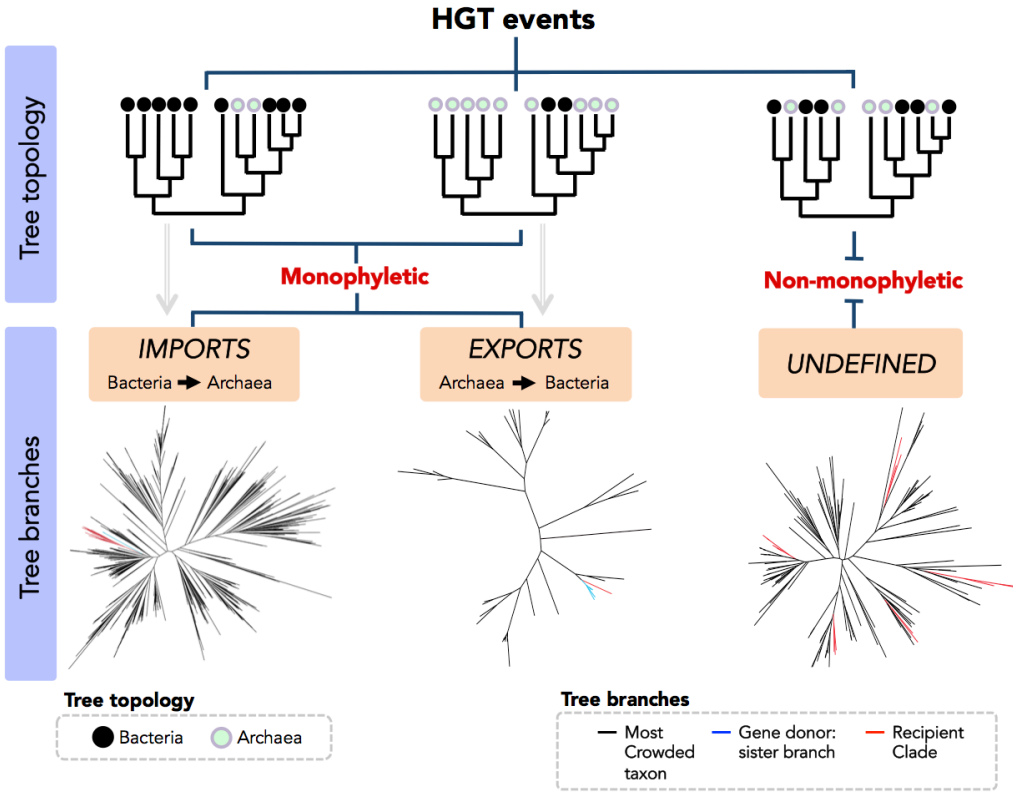


Figure 5.1. HGT detection and classification: branching topology and direction of the transfers. HGTs were identified based on the branching topology (see tree topologies at top) between archaeal and bacterial leaves (bipartition) in a tree previously defined (see Table 5.1 and 5.2), while the direction of the gene transfer was based on the density of species, i.e. out of Africa approach (see tree branches at bottom). *Tree topology*: members of archaea (cyan circles) and bacteria (black circles) are represented in each phylogenetic tree. Single archaeal higher taxa (Chapter 6) and Methanosarcinales taxa (Chapter 7) showing monophyly, as defined previously (see Table 5.1 and 5.2), were the first step to filter all trees and proceed to identify the direction of the transfer. *Tree branches*: Prokaryotic taxa were identified and quantified in each protein family as described in methods. The density of archaeal and bacterial species allows identifying well-polarized cases of gene transfer, such as those from Bacteria-to-Archaea (*imports*), and Archaea-to-Bacteria (*exports*), in which the crowded taxon are from Bacteria and Archaea, respectively. *Undefined transfers* were identified when archaeal taxon was non-monophyletic, as shown at left side.

most represented prokaryotic domain and higher taxon (i.e. highest density) within each protein family: 1) *total number of prokaryotic clades*, 2) *total number of clades by domain* (i.e. Archaea or Bacteria), and 3) *total number of species* (only one single strain) *by clade*. A primary filter to identify well-defined cases (i.e. *polarized gene transfers*) was to take into account only those protein families with ≥3 species members belonging to ≥3 different clades, being one of them the clade of our interest (i.e. archaeal clade). A second approach was used to identify the direction of the transfer when the number of clades in a protein family is equal to 2, i.e. when in a tree there is a clade from bacteria and the other is the archaeal clade. For instance, a tree depicting a HGT event in a protein family tree formed with members belonging to Methanosarcinales and a single bacterial clade only. Here, the direction was calculated according to the density of the species located in every clade. However, these cases might represent low reliability (or without confidence) because their limited taxonomic representation, as has been commented previously (Koonin et al. 2001). See examples in the Figure 5.2.

Table 5.1. Classification of the topologies in the phylogenetic trees for protein families used for the archaeal clades.

HGT in tree	*Topology*	*HGTs group*	*No. arch clades*	*Definition*
No bacterial homologs	-	Archaea specific	Single clade	Families with no bacterial homolog containing only proteins of one single archaeal clade.
	-		≥2 clades	Families with no bacterial homolog containing only proteins of more than two single archaeal clade.
Bacterial homologs	Monophyletic	Import	Single clade	HGT identified in a family containing only proteins of one single archaeal clade, showing monophyly, and bacterial homologs.
		Export	≥1 clades	HGT identified in a family containing only archaeal proteins of one single clade, and only one bacterial phylum.
	No monophyletic	Undefined	≥1 clades	HGT identified in a family containing archaeal proteins of ≥1 archaeal clades, showing no monophyly, but having bacterial homologs.

Table 5.2. Classification of the topologies in the phylogenetic trees for protein families used for the Methanosarcinales study.

HGT in tree	*Topology*	*HGTs group*	*No. arch clades*	*Definition*
Absence of bacteria	-	Msa specific	Single clade	Families with no bacterial homologs, containing only proteins of the Methanosarcinales clade.
		Archaea specific	≥2 clades	Families with no bacterial homologs, containing only proteins of Methanosarcinales and other archaea.
Bacterial homologs	Monophyletic	Import-Msa	Single clade	HGT identified in a family containing only proteins of Methanosarcinales, showing monophyly, and bacterial homologs.
		Import-Archaea	≥2 clades	HGT identified in a family containing proteins of Methanosarcinales and other archaea, showing monophyly, and bacterial homologs.
		Export-Msa	Single clade	HGT identified in a family containing only proteins of Methanosarcinales, and one bacterial phylum.
		Export-Archaea	≥1 clades	HGT identified in a family containing only proteins of Methanosarcinales and other archaea, and one bacterial phylum.
	No monophyletic	Undefined	≥1 clades	HGT identified in a family containing archaeal proteins from ≥1 archaeal clades and bacterial homologs.

In any of the previous cases, we consider HGT transfers from Bacteria to Archaea as "*imports"*, and HGT transfers from Archaea to Bacteria as "*exports"*. Cases represented as non-monophyletic were designed as "*undefined transfers*" (see Figure 5.1). Protein families with no bacterial homologs represent exclusive cases occurring in archaea, therefore such cases were designed as "*archaeal specific*". Specific definitions for these cases for the study of the archaeal clades (Chapter 6) and Methanosarcinales (Chapter 7) are available in Table 5.1 and Table 5.2, respectively.

Distribution of genes in Methanosarcinales protein families. A presence-absence pattern (PAP) analysis was performed using an in-house PERL script for the gene distributions of the archaeal clades (Chapter 6) and Methanosarcinales (Chapter 7) across the protein families. Distributions were plotted separately for the different tree topology categories (see Table 5.1) by using a MATLAB plotting function (vR2013a, 8.1.0.604).

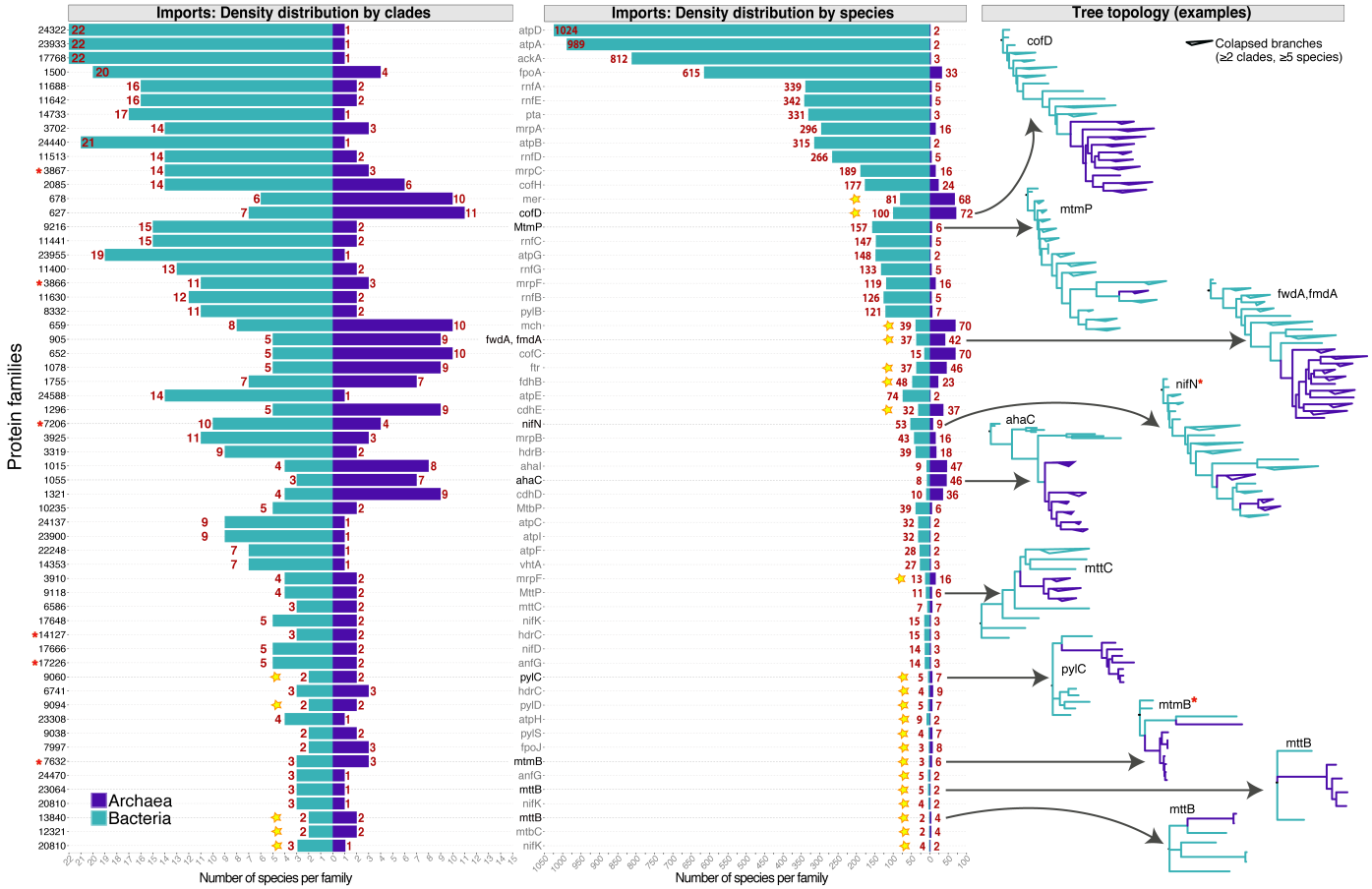


Figure 5.2. Density distribution estimated for protein families identified as imports involved in the methane and energy metabolism. A total of 60 imports were detected by using the *density distribution criterion* calculating the frequency of clades (barplot at left side), and species (barplot at left side) in archaeal protein families (total represented numbers in archaea and bacteria are next to each bar). Cluster number (left y axis) and its corresponding gene names (central y axis) are shown. Examples of the tree topology of some imports are shown (trees at left side). Color of the bars and branches depicted a bacterial (cyan) or archaeal (blue) proteins. Genes with an ambiguous polarized 'density distribution' of the clades—or in case, of the species—, are highlight with a start. Asterisks in red represent cases where ≥2 archaeal clades form a monophyletic group, but split, at least, by ≥1 bacterial proteins.

Table 5.3. Reference genes of transporters and multiple membrane-bound complexes involved in the electron transfer, energy conservation, nitrogen fixation.

Complex name	Clade	Reference organism	Gene name	Pathway	KO number	Gene	EC number
Formate dehydrogenase [1]	Methanococcales	Methanococcus maripaludis S2	*MMP1298*	H	K00123	fdhA	[EC:1.2.1.2]
			MMP0139	H	K00125	fdhB	[EC:1.2.1.2]
			MMP1233	H	K00126	fdhD	[EC:1.2.1.2]
			MMP1301	H	K02379	fdhC	[EC:1.2.1.2]
heterodisulfide reductase [2]	Methanobacteriales	M. thermautotrophicus str. Delta H	*MTH1381*	H	K03388	hdrA	[EC:1.8.98.1]
			MTH1879	H	K03389	hdrB	[EC:1.8.98.1]
			MTH1878	H	K03390	hdrC	[EC:1.8.98.1]
	Methanosarcinales	Methanosarcina acetivorans C2A	*MA0688*	A	K08264	hdrD	[EC:1.8.98.1]
			MA0687	A	K08265	hdrE	[EC:1.8.98.1]
F420-nonreducing hydrogenase [3]	Methanosarcinales	Methanosarcina mazei Go1	*MM_2313*	HIM	--	vhoA	[EC:1.12.2.-]
			MM_2312	HIM	--	vhoC	[EC:1.12.2.-]
			MM_2314	HIM	--	vhoG	[EC:1.12.2.-]
		Methanosarcina acetivorans C2A	*MA1142*	HIM	vhtA	vhtA	[EC:-.-.-.-]
			MA1143	HIM	vhtC	vhtC	[EC:-.-.-.-]
			MA1144	HIM	vhtD	vhtD	[EC:-.-.-.-]
Fpo: F(420)H(2) hydrogenase [4]	Methanosarcinales	Methanosarcina acetivorans C2A	*MA1495*	AIM	K00330	fpoA	[EC:1.6.5.3]
			MA1496	AIM	K00331	fpoB	[EC:1.6.5.3]
			MA1497	AIM	K00332	fpoC	[EC:1.6.5.3]
			MA1498	AIM	K00333	fpoD	[EC:1.6.5.3]
			MA1499	AIM	K00337	fpoH	[EC:1.6.5.3]
			MA1500	AIM	K00338	fpoI	[EC:1.6.5.3]
			MA1501	AIM	K00339	fpoJ	[EC:1.6.5.3]
			MA1503	AIM	K00340	fpoK	[EC:1.6.5.3]
			MA1504	AIM	K00341	fpoL	[EC:1.6.5.3]
			MA1505	AIM	K00342	fpoM	[EC:1.6.5.3]
			MA1506	AIM	K00343	fpoN	[EC:1.6.5.3]
			MA1509	AIM	--	fpoO	[EC:-.-.-.-]
			MA3732	AIM	--	fpoF	[EC:-.-.-.-]
		Methanosaeta concilii GP6	MCON_3103	AIM	K00335	fpoF	[EC:1.6.5.3]
			MCON_3101	AIM	K00334	fpoE	[EC:1.6.5.3]
F-ATPase complex [5]	Methanosarcinales	Methanosarcina acetivorans C2A	MA2440	AIMIH	K02114	atpC	[EC:3.6.3.14]
			MA2441	AIMIH	K02112	atpD	[EC:3.6.3.14]
			MA2433	AIMIH	K02115	atpG	[EC:3.6.3.14]
			MA2434	AIMIH	K02111	atpA	[EC:3.6.3.14]
			MA2438	AIMIH	--	atpH	[EC:3.6.3.14]
			MA2435	AIMIH	K02109	atpF	[EC:3.6.3.14]
			MA2436	AIMIH	K02110	atpE	[EC:3.6.3.14]
			MA2437	AIMIH	K02108	atpB	[EC:3.6.3.14]
			MA2439	AIMIH	K02116	atpI	[EC:3.6.3.14]
A-ATPase complex [6]	Methanosarcinales	Methanosarcina acetivorans C2A	MA4158	AIMIH	K02117	atpA	[EC:3.6.3.14]
			MA4159	AIMIH	K02118	atpB	[EC:3.6.3.14]
			MA4156	AIMIH	K02119	atpC	[EC:3.6.3.14]
			MA4160	AIMIH	K02120	atpD	[EC:3.6.3.14]
			MA4152	AIMIH	K02121	atpE	[EC:3.6.3.14]
			MA4157	AIMIH	K02122	atpF	[EC:3.6.3.14]
			MA4152	AIMIH	K02107	atpH	[EC:3.6.3.14]
			MA4153	AIMIH	K02123	atpI	[EC:3.6.3.14]
			MA4154	AIMIH	K02124	atpK	[EC:3.6.3.14]
Rnf complex [7]	Methanosarcinales	Methanococcoides burtonii DSM 6242	Mbur_1384	A	K03615	rnfC	[EC:-.-.-.-]
			Mbur_1380	A	K03617	rnfA	[EC:-.-.-.-]
			Mbur_1382	A	K03612	rnfG	[EC:-.-.-.-]
			Mbur_1383	A	K03614	rnfD	[EC:-.-.-.-]
			Mbur_1381	A	K03613	rnfE	[EC:-.-.-.-]
			Mbur_1379	A	K03616	rnfB	[EC:-.-.-.-]
			Mbur_1385	A	MA0658	rnfX	[EC:-.-.-.-]
			Mbur_1378	A	MA0665	rnfY	[EC:-.-.-.-]
Ech hydrogenase [8]	Methanosarcinales	Methanosarcina barkeri str. Fusaro	Mbar_A0152	HIA	K14086	echA	[EC:1.18.6.1]
			Mbar_A0151	HIA	K14087	echB	[EC:-.-.-.-]
			Mbar_A0150	HIA	K14088	echC	[EC:-.-.-.-]
			Mbar_A0149	HIA	K14089	echD	[EC:-.-.-.-]
			Mbar_A0148	HIA	K14090	echE	[EC:-.-.-.-]
			Mbar_A0147	HIA	K14091	echF	[EC:-.-.-.-]
Nif nitrogenase [9]	Methanosarcinales	Methanosarcina barkeri str. Fusaro	Mbar_A1552	--	K00531	anfG	[EC:-.-.-.-]
			Mbar_A0171	--	K02588	NifH	[EC:-.-.-.-]
			Mbar_A0168	--	K02586	nifD	[EC:-.-.-.-]
			Mbar_A1551	--	K02591	nifK	[EC:-.-.-.-]
			Mbar_A0165	--	K02592	nifN	[EC:-.-.-.-]
			Mbar_A0166	--	K02587	nifE	[EC:-.-.-.-]
			Mbar_A2279	--	K02590	nifI	[EC:-.-.-.-]
Acetate [10] and methylamines transporters [11]	Methanosarcinales	Methanosarcina acetivorans C2A	MA4008	A	--	aceP	[EC:-.-.-.-]
		Methanosarcina mazei Go1	MM_2046	M	--	MttP	[EC:-.-.-.-]
			MM_2964	M	--	MtbP	[EC:-.-.-.-]
			MM_1434	M	--	MtmP	[EC:-.-.-.-]
Antiporter Na+ / H+ [12]	Methanosarcinales	Methanosarcina acetivorans C2A	MA4572	AIM	K05565	mrpA	[EC:-.-.-.-]
			MA4665	AIM	K05566	mrpB	[EC:-.-.-.-]
			MA4570	AIM	K05567	mrpC	[EC:-.-.-.-]
			MA4569	AIM	K05568	mrpD	[EC:-.-.-.-]
			MA4568	AIM	K05569	mrpE	[EC:-.-.-.-]
			MA4567	AIM	K05570	mrpF	[EC:-.-.-.-]
			MA4566	AIM	K05571	mrpG	[EC:-.-.-.-]
			MA4565	AIM	--	mrpX	[EC:-.-.-.-]
Biosynthesis Pyrrolysine [13]	Methanosarcinales	Methanosarcina acetivorans C2A	MA0154	M	K16180	pylB	[EC:5.4.99.58]
			MA0153	M	K16181	pylC	[EC:-.-.-.-]
			MA0152	M	K16182	pylD	[EC:-.-.-.-]
			MA0155	M	K11627	pylS	[EC:6.1.1.26]
Biosynthesis Cofactor 420 [14]	Methanosarcinales	Methanosarcina acetivorans C2A	MA1488	H	K14941	cofC	[EC:-.-.-.-]
			MA0714	H	K11212	cofD	[EC:-.-.-.-]
			MA0135	H	K12234	cofE	[EC:-.-.-.-]
			MA3512	H	K12234	cofE	[EC:-.-.-.-]
			MA1491	H	K11780	cofG	[EC:-.-.-.-]
			MA1489	H	K11781	cofH	[EC:-.-.-.-]
			Mbur_1304	H	K11781	cofH	[EC:-.-.-.-]

Pathway (header title): Associated methanogenic pathway: H, hydrogenotrophic; A, aceticlastic; M, methylotrophic. Reference genes were obtained from the following studies: [1]Kanehisa and Goto (2000), [2]Galagan et al. (2002), [3,4,5,6,7,8]Rohlin and Gunsalus (2010), Deppenmeier (1995), [6]Deppenmeier (2002), [10]Welte et al. (2014), [11]Hovey et al. (2005), [12]Li et al. (2006), [13]Mahapatra et al. (2006), [14]Wood et al. (2003).

Table 5.4. Reference genes involved in the methanogenesis pathways.

Complex name	Reference organism	Gene name	Pathway	KO number	Gene	EC number
Formylmethanofuran_dehydrogenase	*Methanosarcina mazei Go1*	MM_1566	H	K00200	fwd/fmdA	[EC:1.2.99.5]
		MM_1569	H	K00201	fwd/fmdB	[EC:1.2.99.5]
		MM_1567	H	K00202	fwd/fmdC	[EC:1.2.99.5]
		MM_1568	H	K00203	fwd/fmdD	[EC:1.2.99.5]
		MM_1564	H	K11261	fwd/fmdE	[EC:1.2.99.5]
		MM_1565	H	K00205	fwd/fmdF	[EC:-.-.-.-]
		MM_1824	H	K11260	fwdG	[EC:-.-.-.-]
	Methanococcus maripaludis S2	MMP1244	H	K00204	fwdH	[EC:-.-.-.-]
Formylmethanofuran tetrahydromethanopterin methyltransferase	*Methanosarcina mazei Go1*	MM_1321	H	K00672	ftr	[EC:2.3.1.101]
Methenyltetrahydromethanopterin cyclohydrolase	*Methanosarcina mazei Go1*	MM_2653	H	K01499	mch	[EC:3.5.4.27]
Methylenetetrahydromethanopterin dehydrogenase	*Methanosarcina mazei Go1*	MM_1108	H	K00319	mtd	[EC:1.5.99.9/98.1]
Methenyltetrahydromethanopterin hydrogenase	*Methanothermobacter marburgensis*	MMP0127	H	K13942	hmd	[EC:1.12.98.2]
Cofactor F420 reducing hydrogenase	*Methanosarcina mazei Go1*	MM_3045	H	K00440	frhA	[EC:1.12.98.1]
		MM_3042	H	K00441	frhB	[EC:1.12.98.1]
		MM_3044	H	K00442	frhD	[EC:1.12.98.1]
		MM_3043	H	K00443	frhG	[EC:1.12.98.1]
Methylenetetrahydromethanopterin reductase	*Methanosarcina mazei Go1*	MM_0628	H	K00320	mer	[EC:1.5.98.2]
Acetil CoA synthetase	*Methanosaeta thermophila PT*	MTHE_1194	A	K01895	ACS	[EC:6.2.1.1]
Acetate_kinase	*Methanosarcina acetivorans C2A*	MA3606	A	K00925	ackA	[EC:2.7.2.1]
Phosphate acetyltransferase	*Methanosarcina acetivorans C2A*	MA3607	A	K00625	pta	[EC:2.3.1.8]
CODH ACDS	*Methanosarcina acetivorans C2A*	MA3860	A	K00192	cdhA	[EC:1.2.7.4]
		MA3861	A	K00195	cdhB	[EC:-.-.-.-]
		MA3862	A	K00193	cdhC	[EC:2.3.1.-]
		MA3864	A	K00194	cdhD	[EC:2.1.1.245]
		MA3865	A	K00197	cdhE	[EC:2.1.1.245]
Dimethylamine	*Methanosarcina mazei Go1*	MM_2051	M	K16178	mtbB	[EC:2.1.1.249]
		MM_2052	M	K16179	mtbC	[EC:-.-.-.-]
Monomethylamine	*Methanosarcina mazei Go1*	MM_3335	M	K16176	mtmB	[EC:2.1.1.248]
		MM_3334	M	K16177	mtmC	[EC:-.-.-.-]
Trimethylamine	*Methanosarcina mazei Go1*	MM_2047	M	K14084	mttC	[EC:-.-.-.-]
		MM_2048	M	K14083	mttB	[EC:2.1.1.250]
Core methyamine	*Methanosarcina mazei Go1*	MM_1439	M	K14082	mtbA	[EC:2.1.1.247]
Methanol	*Methanosarcina acetivorans C2A*	MA0779	M	K14080	mtaA	[EC:2.1.1.246]
		MA0455	M	K04480	mtaB1	[EC:2.1.1.90]
		MA0456	M	K14081	mtaC	[EC:-.-.-.-]
Dimethylsulfide	*Methanosarcina mazei Go1*	MM_2428	M	K16955	mtsB	[EC:-.-.-.-]
		MM_2427	M	K16954	mtsA	[EC:-.-.-.-]
N-5 Methyltetrahydromethanopterin: coenzyme M methyltransferase	*Methanosarcina mazei Go1*	MM_1543	H|A|M	K00577	mtrA	[EC:2.1.1.86]
		MM_1544	H|A|M	K00578	mtrB	[EC:2.1.1.86]
		MM_1545	H|A|M	K00579	mtrC	[EC:2.1.1.86]
		MM_1546	H|A|M	K00580	mtrD	[EC:2.1.1.86]
		MM_1547	H|A|M	K00581	mtrE	[EC:2.1.1.86]
		MM_1542	H|A|M	K00582	mtrF	[EC:2.1.1.86]
		MM_1541	H|A|M	K00583	mtrG	[EC:2.1.1.86]
		MM_1540	H|A|M	K00584	mtrH	[EC:2.1.1.86]
Methyl-coenzyme_M_reductase	*Methanothermobacter thermautotrophicus str. Delta H*	MM_1240	H|A|M	K00399	mcrA	[EC:2.8.4.1]
		MM_1244	H|A|M	K00401	mcrB	[EC:2.8.4.1]
		MM_1241	H|A|M	K00402	mcrG	[EC:2.8.4.1]
		MM_1242	H|A|M	K03421	mcrC	[EC:-.-.-.-]
		MM_1243	H|A|M	K03422	mcrD	[EC:-.-.-.-]

Pathway header title: Methanogenesis: H, hydrogenotrophic methanogenesis; A, aceticlastic methanogenesis; M, methylotrophic methanogenesis. Reference genes were obtained from the following studies: Kanehisa and Goto (2000), Galagan et al. (2002), Hovey et al. (2005), Soares et al. (2005), Rohlin and Gunsalus (2010), Berger et al. (2012).

Detection of functional categorization of genes and pathways. A search of functional classification for all 6,103,411 genes in 1,981 prokaryotic genomes was performed by using the Cluster of Orthologous Groups (COG) database (Tatusov 2003) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (v07.2015) (Kanehisa and Goto 2000). Putative roles of genes in multiple metabolic pathways were assigned via the KEGG Orthology (KO) system. Metabolic pathways referred in the KO system database were downloaded from the KEGG website. Thus, the most frequent functional annotation for each of the protein families from the COG and KEGG databases was taken as the one to describe best a particular protein family.

Table 5.5. Reference genes involved in osmotic stress in salt environments.

Physiological role	Main function	Description	Gene ID
NaCl-induced genes	Glycine betaine Transport (Ota) (Roeßler et al., 2002)	Glycine betaine transporter, ATP-binding protein (OtaA)	MM_0040
		Glycine betaine transporter, permease (OtaB)	MM_0041
		Glycine betaine transporter, substrate-binding protein (OtaC)	MM_0042
	Biosynthesis of Ne-acetyl-b-lysine (Abl) (Pfluger et al., 2003)	Lysine-2,3-aminomutase (AblA)	MM_0934
		b-Lysine acetyltransferase (AblB)	MM_0935
	Transport and binding proteins: ions, amino acids, and phosphate	Amino acid permease	MM_0530
		Phosphate-binding protein	MM_0578
		Cobalt transporter, ATP-binding protein	MM_0887
		ABC transporter, ATP-binding protein	MM_0950
		Phosphate transporter, phosphate-binding protein (PstS)	MM_2005
		Phosphate transporter, permease (PstC)	MM_2006
		Phosphate transporter, ATP-binding protein (PstB)	MM_2009
		Hypothetical protein	MM_2637
		Sodium/glutamate symport carrier protein (GltS)	MM_1782
	Cell surface modifying enzymes	Conserved protein	MM_0353
		Polysaccharide deacetylase	MM_1117
		Glycosyl transferase	MM_1121
		Surface layer protein B (SlpB)	MM_1589
		Hypothetical protein	MM_1897
		Integral membrane protein	MM_1933
		Hypothetical protein	MM_2529
		Surface layer protein (putative)	MM_2587
	Stress response	Putative heat shock protein	MM_1254
		Universal stress protein	MM_1452
	Regulatory functions	Hypothetical protein	MM_0077
		Conserved protein	MM_1605
		Hypothetical sensory transduction histidine kinase	MM_1671
		Conserved protein	MM_1882
	Na1 exporter	Putative ATP-dependent Na1 efflux pump (NatB)	MM_1056
	Metabolic enzymes and facilitating proteins	Thiamine biosynthesis protein	MM_0182
		Oxidoreductase (flavoprotein)	MM_0718
		Conserved protein	MM_0883
		Heterodisulfide reductase, subunit HDRB	MM_0980
		Conserved protein	MM_1072
		Hypothetical protein	MM_1103
		Coenzyme F420 hydrogenase, b subunit	MM_1225
		FdhD protein	MM_1506
		Iron–sulfur flavoprotein	MM_1585
		Deoxycytidine triphosphate deaminase	MM_1628
		Malate dehydrogenase	MM_1966
		Conserved protein	MM_2018
		Flavodoxin	MM_3140
		Coenzyme F420 hydrogenase, b subunit	MM_3264
	DNA modifying proteins	Putative flagella-related protein H	MM_0939
		ATP-dependent DNA helicase	MM_1217
		Type I restriction-modification system, restriction subunit	MM_1661
	Protein fate: degradation, secretion, and stabilization	Hypothetical protein	MM_1999
		Protease HTPX homolog	MM_2115
		Sec-independent protein translocase	MM_2585
	Variety of known/putative proteins	HTH DNA-binding protein	MM_0444
		Iron–sulfur cluster-binding protein	MM_1019
		Zinc metalloprotease	MM_1523
		Phosphohydrolase	MM_1873
		Histone actyltransferase (ELP3 family)	MM_1974
		Putative serine/threonine protein kinase	MM_2122
		Transposase	MM_3228
	Hypothetical/conserved proteins	Hypothetical protein	MM_0970
		Hypothetical protein	MM_0971
		Conserved protein	MM_1067
		Conserved protein	MM_1071
		Conserved protein	MM_1097
		Conserved protein	MM_1118
		Conserved protein	MM_1120
		Conserved protein	MM_1129
		Hypothetical protein	MM_1409
		Putative nucleotidyltransferase	MM_1491
		Conserved protein	MM_1520
		Molybdopterin biosynthesis MoeB protein	MM_1533
		Putative metal dependent hydrolase	MM_1555
		Conserved protein	MM_1574
		Conserved protein	MM_1713
		Hypothetical protein	MM_1779
		Hypothetical protein	MM_1784
		Conserved protein	MM_1872
		Hypothetical protein	MM_1914
		Hypothetical protein	MM_1929
		Hypothetical protein	MM_1964
		Hypothetical cytosolic protein	MM_1984
		Hypothetical protein	MM_2043
		Conserved transmembrane protein	MM_2378
		Conserved protein	MM_2882
		Hypothetical protein	MM_3106
		Hypothetical protein	MM_3111
		Hypothetical protein	MM_3162
NaCl-repressed genes	ATPase	A1A0 H1 ATPase, subunit K	MM_0784
	Metabolic enzymes and facilitating proteins	Glucose-1-phosphate thymidylyltransferas	MM_0299
		Sulfite reductase, assimilatory-type	MM_0360
		N-acetyl-g-glutamyl-phosphate reductase	MM_0476
		Coenzyme F420-dep. methylene 4HMPT reductase	MM_0628
		b-Ketoacyl synthase/thiolase	MM_0870
		Hydroxymethylglutaryl-CoA synthase	MM_0871
		Pyruvate synthase, g subunit	MM_1342
	Transport and binding proteins	Dipeptide ABC transporter, binding protein	MM_0001
		Trk system potassium uptake protein (TrkA)	MM_2503

Detection of HGT events in the most versatile biological processes of Methanosarcinales. We further estimated the presence of HGT events in specific metabolic functions in Methanosarcinales including: **1) *energy metabolism***: protein complexes of membrane-bound electron transport and energy conservation system, as well as ***nitrogen fixation*** performed by the Nif protein complex**, 2) *methanogenesis***: proteins playing a role in the core reactions of hydrogenotrophic, aceticlastic and methylotrophic methanogenesis, **3) *salt adaptation***: proteins involved in the secondary response (*salt-out*) to the osmotic pressure in Methanosarcinales.

To that end, detailed information for a set of 188 *reference genes* was compiled from experimental and bioinformatics literature. After processing this information, the final set of 'reference genes' encompass 84 reference genes representing the electron transport and energy conservation, methanogenesis, and nitrogen fixation processes was selected for eight specific methanogens (see Table 5.3 and Table 5.4). Furthermore, 94 genes experimentally reported in *Methanosarcina mazei* (Pflüger et al. 2007) were included as "reference proteins" for osmotic pressure in salt adaptation (see Table 5.5). Methanogens containing the reference genes belong to *Methanococcales* (*M. maripaludis*) and *Methanosarcinales*, including genera such as *Methanosaeta*, *Methanosarcina* and *Methanococcoides*. Also, the genes of the methyl coenzyme M reductase (mcr) from *Methanobacter thermoautotrophicus* were used as reference.

From the set of 188 reference genes, 93 genes are involved in membrane protein complexes (membrane-bound complexes), cytosolic protein complexes and single membrane transporters (Table 5.3). The latest ones are proteins that drive the uptake of the following substrates: acetate (*aceP*), mono-methylamine (*mtmP*), di-methylamine (*mtbP*), tri-methylamine (*mttP*), and formate (*fdhC*), and the antiporter *Mrp* of Na^+/H^+. Other protein complexes are also included, such as *Ech*, *Hdr*, *Vho*, *Vht*, *Fpo*, *Rnf, F-ATPase* (a.k.a. *atp*), A-ATPase (a.k.a. *aha*) and *Nif*. The remaining 11 reference genes represent the set of genes (*operon*) related to the biosynthesis of pyrrolysine (*pylSBCD*) and F_{420} (*cofC*, *cofD*, *cofE*, *cofG*, *cofH*).

The remaining set of 84 'reference genes' represents the methanogenesis metabolism, and include the core reactions of the hydrogenotrophic, aceticlastic, and the methylotrophic pathways, which comprise 27, 18 and 23 genes, respectively. In addition to this, 16 genes from the *mtr* and *mcr* complexes were included in the set of reference genes (see Table 5.4). Finally, all reference genes were mapped across the 5,033 Methanosarcinales protein families to identify: 1) the archaeal and bacterial homologs, 2) the presence or absence of HGT events using the annotation of defined tree topologies, and 3) to define the origin of the putative HGT events (see Table 5.2).

Detection of HGT events in the system of membrane complexes involved in the aerobic respiration of Halobacteria. The acquisition of membrane respiratory complexes via HGT was evaluated in the context of 23 Halobacteriales genomes. The corresponding membrane respiratory complexes that were included in this analysis are: *NADH dehydrogenase* (complex I), *Succinate dehydrogenase* (complex II), *cytochrome* bc_1 (complex III), 4) *cytochrome c oxidase* (complex IV), and 5) *ATPase* (complex V). The corresponding reference gene annotation of this respiratory system was identified with the KO orthology system annotation from the KEGG database (see Table 5.6).

Overall, a collection of 44 KO orthology annotations was associated to each component of the aerobic respiration (see Table 5.6). From this, 14 KEGG annotations correspond to complex I (*nuoA-N*), 4 annotations to complex II (*sdhABCD*), 10 annotations to complex III (Rieske iron

Table 5.6. Reference KO number annotations from KEGG database to identify genes involved in aerobic respiration.

Complex no.	Complex name	No. Genes	KO number	Gene name	Complex name	EC ID	Domain*
Complex I	*NADH dehydrogenase [NADH]*	1	K00330	nuoA	NADH-quinone oxidoreductase subunit A	[EC:1.6.5.3]	A, B
		2	K00331	nuoB	NADH-quinone oxidoreductase subunit B	[EC:1.6.5.3]	A, B
		3	K00332	nuoC	NADH-quinone oxidoreductase subunit C	[EC:1.6.5.3]	A, B
		4	K13378	nuoCD	NADH-quinone oxidoreductase subunit C/D	[EC:1.6.5.3]	A, B
		5	K00334	nuoE	NADH-quinone oxidoreductase subunit E	[EC:1.6.5.3]	A, B
		6	K00335	nuoF	NADH-quinone oxidoreductase subunit F	[EC:1.6.5.3]	A, B
		7	K00336	nuoG	NADH-quinone oxidoreductase subunit G	[EC:1.6.5.3]	A, B
		8	K00337	nuoH	NADH-quinone oxidoreductase subunit H	[EC:1.6.5.3]	A, B
		9	K00338	nuoI	NADH-quinone oxidoreductase subunit I	[EC:1.6.5.3]	A, B
		10	K00339	nuoJ	NADH-quinone oxidoreductase subunit J	[EC:1.6.5.3]	A, B
		11	K00340	nuoK	NADH-quinone oxidoreductase subunit K	[EC:1.6.5.3]	A, B
		12	K00341	nuoL	NADH-quinone oxidoreductase subunit L	[EC:1.6.5.3]	A, B
		13	K00342	nuoM	NADH-quinone oxidoreductase subunit M	[EC:1.6.5.3]	A, B
		14	K00343	nuoN	NADH-quinone oxidoreductase subunit N	[EC:1.6.5.3]	A, B
Complex II	*Succinate dehydrogenase*	1	K00239	sdhA	succinate dehydrogenase flavoprotein subunit	[EC:1.3.99.1]	A, B
		2	K00240	sdhB	succinate dehydrogenase iron-sulfur subunit	[EC:1.3.99.1]	A, B
		3	K00241	sdhC	succinate dehydrogenase cytochrome b556 subunit	[EC:-.-.-.-]	A, B
		4	K00242	sdhD	succinate dehydrogenase membrane anchor subunit	[EC:-.-.-.-]	A, B
Complex III	*Cytochrome C reductase: menaquinol*	1	K03887	qcrB,petB	menaquinol-cytochrome c reductase cytochrome b subunit	[EC:-.-.-.-]	A, B, E
		2	K03888	qcrC,petD	menaquinol-cytochrome c reductase cytochrome b/c subunit	[EC:-.-.-.-]	A, B, E
		3	K03886	qcrA,petC	menaquinol-cytochrome c reductase iron-sulfur subunit	[EC:1.10.2.-]	A, B, E
		4	K00412	CytB,petB	ubiquinol-cytochrome c reductase cytochrome b subunit	[EC:-.-.-.-]	A, B, E
		5	K03891	qcrB	ubiquinol-cytochrome c reductase cytochrome b subunit	[EC:-.-.-.-]	A, B, E
	Cytochrome C reductase: ubiquinol	6	K00410	fbcH	ubiquinol-cytochrome c reductase cytochrome b/c1 subunit	[EC:-.-.-.-]	A, B, E
		7	K03889	qcrC	ubiquinol-cytochrome c reductase cytochrome c subunit	[EC:-.-.-.-]	A, B, E
		8	K00413	Cyt1,petC	ubiquinol-cytochrome c reductase cytochrome c1 subunit	[EC:-.-.-.-]	A, B, E
		9	K00411	RIP1,petA	ubiquinol-cytochrome c reductase iron-sulfur subunit	[EC:1.10.2.2]	A, B, E
		10	K03890	qcrA	ubiquinol-cytochrome c reductase iron-sulfur subunit	[EC:-.-.-.-]	A, B, E
Complex IV	*Cytochrome C oxidase*	1	K02277	coxD,ctaF	cytochrome c oxidase subunit IV	[EC:1.9.3.1]	A, B
		2	K02276	coxC	cytochrome c oxidase subunit III	[EC:1.9.3.1]	A, B
		3	K02274	coxA	cytochrome c oxidase subunit I	[EC:1.9.3.1]	A, B
		4	K15408	coxAC	cytochrome c oxidase subunit I+III	[EC:1.9.3.1]	A, B
		5	K02275	coxB	cytochrome c oxidase subunit II	[EC:1.9.3.1]	A, B
	Cytochrome BD complex	6	K00425	cydA	cytochrome d ubiquinol oxidase subunit I	[EC:1.10.3.-]	A, B
		7	K00426	cydB	cytochrome d ubiquinol oxidase subunit II	[EC:1.10.3.-]	A, B
Complex V	*V/A-type APTase*	1	K02117	ahaA	V-type H+-transporting ATPase subunit A	[EC:3.6.3.14]	A, B
		2	K02123	ahaI	V-type H+-transporting ATPase subunit I	[EC:3.6.3.14]	A, B
		3	K02118	ahaB	V-type H+-transporting ATPase subunit B	[EC:3.6.3.14]	A, B
		4	K02124	ahaK	V-type H+-transporting ATPase subunit K	[EC:3.6.3.14]	A, B
		5	K02119	ahaC	V-type H+-transporting ATPase subunit C	[EC:3.6.3.14]	A, B
		6	K02120	ahaD	V-type H+-transporting ATPase subunit D	[EC:3.6.3.14]	A, B
		7	K02121	ahaE	V-type H+-transporting ATPase subunit E	[EC:3.6.3.14]	A, B
		8	K02122	ahaF	V-type H+-transporting ATPase subunit F	[EC:3.6.3.14]	A, B
		9	K02107	ahaH	V/A-type H+-transporting ATPase subunit G/H	[EC:-.-.-.-]	A, B

* Domain: A, Archaea; B, Bacteria; E, Eukaryotes

sulfur protein, *b*-type cytochromes, and *c*-type cytochrome), 7 annotations to complex IV (5 for *coxBAC* and 2 for *cydABCD*), and 9 annotations correspond to complex V (*atpA-H*). All 44 KO orthologous gene annotation involved in the aerobic respiration were screen across all 25,762 protein families initially reconstructed (Chapter 6). HGT events were identified based on the classification previously created to analyze the origin of archaeal clades via HGTs (see Table 5.1). Genes with no HGTs also were counted in this analysis.

Detection of the operon-like gene organization of the respiratory chain in Halobacteria. The putative operons or operon-like organization of a set of genes was detected by using the information of the positions of the genes (from the NCBI gene annotation), the KEGG gene annotation and literature survey. Once the genes were physically located along the genome, a calculated distance between all gene pairs forming a cluster was performed through an in-house PERL script. The schematic overview and distributions of the cluster of genes within, and across several archaeal genomes, was performed using an in-house R script. Graphical processing was performed with the genoPlotR v0.8.3 (Lionel et al. 2010) and ggbio v3.2 (Yin et al. 2012) R packages.

Briefly, the search for a putative operon-like organization was performed with a combination of manual curation and standard programming under three simple criteria: **1)** identifying all reference **adjacent gene pairs** on the same strand that belong to a specific membrane bound complex, **2) identifying a minimum number of subunits** that form a specific protein complex; such minimum number was assigned according to each case particularly. For instance, for a complex with three subunits such as sdhA, sdhB, sdhC, the minimum number of adjacent genes is two, whereas for long complexes —such as nuoA-H with 14 subunits—the minimum number of subunits to be found is fixed to the half. Importantly, an intergenic distance (I_d) threshold (I_{dT}) between adjacent genes was not defined. Although a minimum average distance between adjacent genes is expected to be conserved in many cases, we are aware that I_d distances are expected to be highly variable in large genomes because there is a strong correlation between genome size and I_d (Nuñez et al. 2013). Likewise, Halobacteriales genomes are the longest in Archaea, and they do not possess a conservative genome size within the group. So that, **insertions were not a restriction between adjacent genes**, because exists the probability that different genes can be found between adjacent genes (Price et al. 2006) or that a set of clustered genes from an operon (partial or complete) can be transferred (Yap et al. 1999).

Assessing the functionality of proteins acquired through HGT events. In order to identify putative functional domains across all prokaryotic proteins involved in HGTs, two different databases of protein domains were used: PFAM v28.0, a database of protein domain families based on sequence conservation (Finn et al. 2014), and Superfamily v1.75, a database of protein domain families based on structural and functional annotation (Gough et al. 2001). HMMER v3.1b2 was used to perform the search analysis for protein domains profiles against all prokaryotic genomes. A *cut_ga* parameter (cutoff threshold for sequence and domain), which measures the significance of a hit, was used as a threshold to evaluate the conservation of each PFAM domain profile (Punta et al. 2012).

By using in-house PERL scripts, a basic presence/absence profile of the protein domain conservation based on PFAM sequence-based domains between a reference gene and its corresponding xenologous protein, were used to evaluate the functionality of the xenologous proteins in a specific metabolic pathway. For example, if a xenologous protein conserves the same number and type of domains that a reference gene, then such xenologous gene is considered functional or at least with a higher probability to conserve its original functional role. Thus, this approach also helped us to identify xenologous proteins that have being subject to major mutational that might process compromising their original functionality, given that they either do not have any or conserve only part of the domains in comparison to the reference gene. Finally, the graphical representation of such compared distributions was obtained combining both PERL and R in-house scripts.

Finding orthologous genes involved in the methane and energy metabolism among new Methanosarcinales genomes. First, in order to expand the number of Methanosarcinales genomes to performed further comparative analysis, and because the number of Methanosarcinales genomes was limited to 10 at the begging of this thesis (prokaryotic genomes retrieved in July 2012), additional 24 new sequenced Methanosarcinales genomes were downloaded from the NCBI (version of June 2017) (Altschul et al. 1997). Thus, 34 genomes from Methanosarcinales were used for additional genomic comparative analysis. The taxonomy classification for all genomes was obtained from the NCBI taxonomy website (https://www.ncbi.nlm.nih.gov/taxonomy). Second, the functional gene annotation of the new genomes was obtained from the KEGG Orthology (KO) system of the KEGG database (v06.2017) (Kanehisa and Goto 2000).

In order to identify orthologous genes within all Methanosarcinales genomes, a BLAST all-against-all was carried out for all 105,556 proteins. Putative orthologous genes from BLAST were considered after being filtered through the reciprocal best BLAST hit (rBBH) approach using an in-house PERL script with the following parameters: local identity ≥25%, an e-value ≤$1x10^{-10}$, and ≥70% of sequence *coverage*. A total of 2,420,641 rBBH fulfilled the previous conditions, and were thus submitted to a pairwise global identity, as further improvement in the quality of the sequence alignment (≥25% identity), through the Needleman-Wunsch algorithm from the EMBOSS package (version 6.6.0.0) (Rice et al. 2000).

Next, 188 reference genes representing the methane and energy metabolisms, as described earlier in this section, were screened across all candidates of orthologous genes (see Table 5.3 and Table 5.4). A final set of 2,456 orthologous genes were obtained for the 34 Methanosarcinales genomes, and their *functional protein domains* were identified using PFAM v28.0 as described in a previous section (see section "Assessing the functionality of proteins acquired through HGT events").

Growth of Methanosarcinales species under different methanogenic conditions obtained from the literature. Information obtained from experimental reports was consulted among 48 reports for 48 Methanosarcinales species and strains. The methanogenic conditions described here were based on the type of methanogenic pathway performed by each species, hydrogenotrophic, aceticlastic and methylotrophic. The Table 5.7 summarizes this information.

Table 5.7. Experimental reports analyzing the capability of Methanosarcinales species and strains to grow on different methanogenic conditions obtained from the literature.

Group	Species name	C	F	A	M	D	T	D	Mt	Reference
Methanosaeta	*Methanosaeta harundinacea* 6Ac	-	-	+	-	-	-	-	-	Ma et al., 2006
	Methanosaeta thermophila PT	-	-	+	-	-	-	-	-	Kamagata et al., 1992
	Methanosaeta concilii GP6	-	-	+	-	-	-	-	-	Patel and Sprott, 1990
Methanosarcina	*Methanosarcina barkeri str Fusaro*	+	x	+	+	+	+	x	+	Bock & Schönheit 1995; Rosenber et al., 2014
	Methanosarcina barkeri MS	+	x	-	+	+	+	x	+	Bryant & Boone, 1987; Maestrojuán & Boone, 1991
	Methanosarcina barkeri Wiesmoor	x	x	x	x	x	x	x	x	*No reported*
	Methanosarcina barkeri 227	x	x	x	x	x	x	x	x	*No reported*
	Methanosarcina barkeri 3	x	x	x	x	x	x	x	x	*No reported*
	Methanosarcina mazei Go1	-	x	+	x	+	+	x	+	Hovey et al., 2005
	Methanosarcina mazei S6	+/-	x	+/-	+	+	+	x	+	Mah & Kuhn, 1984; Maestrojuán & Boone, 1991
	Methanosarcina mazei Tuc01	x	x	x	x	x	x	x	x	No reported
	Methanosarcina mazei C16	x	x	x	x	x	-	-	-	Maestrojuan et al., 1992
	Methanosarcina vacuolata Z-761	+	-	+	+	+	+	+	+	Zhilina & Zavarzin, 1987; Maestrojuán & Boone, 1991
	Methanosarcina horonobensis	-	-	+	-	+	+	+	+	Shimizu et al., 2011
	Methanosarcina thermophila TM1	+/-	-	+	+	+	+	+	+	Zinder et al., 1985; Zinder & Mah, 1979
	Methanosarcina thermophila CHTI55	x	x	x	x	x	x	x	x	*No reported*
	Methanosarcina sp Kolksee	x	x	x	x	x	x	x	x	*No reported*
	Methanosarcina lacustris	-	-	-	+	+	+	-	+	Simankova et al., 2001
	Methanosarcina acetivorans C2A	-	-	+	+	+	+	+	+	Sowers et al., 1984
	Methanosarcina baltica	-	+/-	+	+	+	+	-	+	von Klein et al., 2002
	Methanosarcina semesiae	-	-	-	+	+	+	+	+	Lyimo et al., 2000
	Methanosarcina soligelidi	+/-	-	+	-	-	-	-	+	Wagner et al., 2013
	Methanosarcina siciliae C2J	-	x	-	x	x	+	+	+	Elberson & Sowers, 1997
	Methanosarcina siciliae T4M	-	-	+/-	x	x	+	+	+	Ni & Boone, 1991; Ni et al., 1994
	Methanosarcina sp MTP4	x	x	x	x	x	x	x	x	*No reported*
	Methanosarcina sp WH1	x	x	x	x	x	x	x	x	*No reported*
	Methanosarcina sp WWM596	x	x	x	x	x	x	x	x	*No reported*
	Methanosarcina siciliae HI350	x	x	x	x	x	+	+	+	Ni et al., 1994
Others	*Methanolobus psychrophilus*	-	-	-	+	+	+	+	+	Zhang et al., 2008
	Methanolobus tindarius	x	x	x	+	+	+	-	+	Konig & Stetter, 1982
	Methanolobus bombayensis	-	-	-	+	+	+	+	+	Kadam et al., 1994
	Methanolobus oregonensis	-	-	-	+	x	x	+	+	Liu & Boone, 1990; Boone 2001
	Methanolobus profundi	-	-	-	+	+	+	-	+	Mochimaru et al., 2009
	Methanolobus taylorii	-	-	-	+	+	+	-	+	Oremland & Boone, 1994
	Methanolobus vulcani	-	-	-	+	+	+	-	+	L'Haridon et al., 2014
	Methanolobus zinderi	-	x	-	+	+	+	x	+	Doerfert et al., 2009
	Methanohalophilus mahii DSM 5219	-	-	-	+	+	+	x	+	Paterek & Smith, 1985, 1988
	Methanohalophilus halophilus	-	-	-	+	+	+	x	+	Wilharm et al., 1991
	Methanohalophilus portucalensis	-	-	-	+	+	+	-	+	Boone et al., 1993
	Methanococcoides burtonii DSM 6242	-	-	-	+	+	+	-	+	Allen et al., 2009
	Methanococcoides methylutens	-	-	-	+	+	+	+	+	Franzmann et al., 1992
	Methanococcoides alaskense	-	-	-	x	x	+	-	-	Singh et al., 2005
	Methanomethylovorans hollandica DMS1	-	-	x	+	+	+	+	+	Lomans et al., 1999
	Methanomethylovorans thermophila	-	-	-	+	+	+	-	+	Jiang et al., 2005
	Methanohalobium evestigatum Z-7303	-	-	-	+	+	+	x	+	Zhilina & Zavarzin, 1987; Zhilina & Svetlichnaya, 1989
	Halomethanococcus doii	-	-	-	+	+	+	x	+	Yu & Kawamura, 1987
	Mdthanimicrococcus blatticola	-	-	-	+	+	+	x	+	Sprenger et al., 2000
	Methanosalsum zhilinae DSM 4017	-	-	-	+	+	+	+	+	Boone & Baker, 2001

+, growth; -, not growth; +/-, growth in some strains; x, no reported or no tested.

Hydrogenotrophic: C, CO2/H2; F, formate. **Aceticlastic:** A, acetate.

Methylotrophic: M, monomethylamine; D, dimethylamine; T, trimethylamine; D, dimethylsulfide; Mt, methanol

6

ORIGINS OF MAJOR ARCHAEAL CLADES CORRESPOND TO BACTERIAL ACQUISITIONS

The origin of species or major prokaryotic groups is subject to different mechanisms, where the HGT influence on these lineage-specific processes has not been explored. In order to study the origin of the archaeal taxonomic clades in the context of their putative bacterial homologs, an extensive phylogenetic analysis was performed to evaluate the effect of the HGT in these major archaeal groups.

First, in order to obtain an evolutionary reference tree for all 134 archaeal genomes, a phylogenetic tree was reconstructed from a weighted concatenation alignment of 70 archaeal single copy genes. Initially, archaeal genomes were subdivided into 15 clades based on their taxonomic *Order* rank; however, because Korarchaeota, Thaumarchaeota and Nanoarchaeota have only 1 and 2 members, they were grouped in a single group. Most of the archaeal groups were arranged according to Makarova et al. (2010) and Nelson-Sathi et al. (2012). Thus, archaeal genomes were organized in 13 groups or *higher taxa* (see phylogenetic tree at left side in Figure 6.1). Second, functional COG category of these universal genes in Archaea shows that most of the functions are used for '*informational processes'*, and with a less amount are those related to '*cellular and membrane processes*'. The 70 universal genes used to reconstruct the unrooted reference tree are *ansA, arf1, cca, dnaG, drg, eif2A, eif5a, eif2G, flpA, gate, glyS, gcp, gsp-3, hisS, infB, map, metG, pth, proS, pheS, pheT, recA, rpsJ, rpsK, rps15p, rpsQ, rps19e, rpsB, rps28e, rpsD, rps4e, rpsE, rps7, rpsH, rpl, rpl15, rpsC, rplP, rpl18p, rplR, rplK, rplU, rl22, rpl24, rplW, rpl30P, rplC, rpl4lp, rplE, rpl7ae, rplB, rpsM, rpsH, rplF, rpsS, rpsI,rimM, rli, rpoE, rpoA, rpoB, rnhB, rfcL, rnz, rio1, rplA, SecY, trm, valS, yyaF.*

6.1 Large number of putative homologous genes between archaea and bacteria shows a massive amount of bacterial genes shared with specific archaeal clades.

Figure 6.1 shows a total of 1,240,079 putative protein homologs between Archaea and Bacteria. The distribution of shared genes in both domains is widely distributed across all members of Bacteria. In Archaea, such homologs proteins are also distributed over all 13 groups, but with a high marked amount of them in members of the Euryarchaeota phylum. Although some Crenarchaeota groups, such as *Sulfolobales* and *Thermoplasmatales,* present a high number of putative homologs (≤300 proteins) with bacterial genomes, the largest number of homologous proteins shared with Bacteria belongs to two specific archaeal groups from Euryarchaeota,

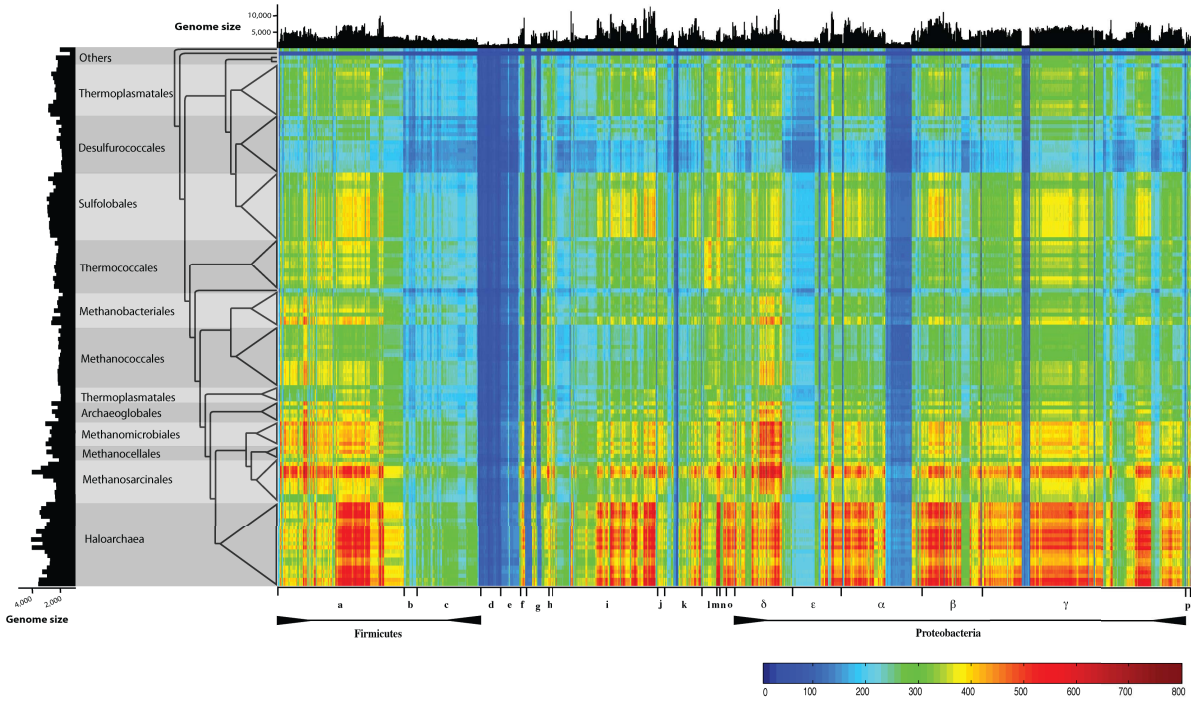


Figure 6.1. Shared gene network between archaeal and bacterial species. Archaeal species were organized into 13 archaeal groups and sorted by the reference tree (left side). In those cases where eubacterial homologs are present (at the bottom), the distribution of bacterial correspond to the phylum (or class) of different groups: *a* = Clostridia*; b* = Erysipelotrichi, Negativicutes; *c* = Bacilli; *d* = Firmicutes; *e* = Chlamydia; *f* = Verrucomicrobia, Planctomycete; *g* = Spirochaete; *h* = Gemmatimonadetes, Synergisteles, Elusimicrobia, Dyctyoglomi, Nitrospirae; *i* =Actinobacteria ; *j* = Fibrobacter, Chlorobi; *k* = Bacteroidetes; *l* = Fusobacteria; Thermatogae, Aquificae, Chloroflexi; *m* = Deinococcus-Thermus; *n* = Cyanobacteria; *o* = Acidobacteria; d, e, a, b, g = Delta, Epsilon, Alpha, Beta and Gamma proteobacteria; *p* = Thermosulfurobateria, Caldiserica, Chysiogenete, Ignavibacteria. Archaeal and bacterial genome sizes (as the total number of proteins) are indicated at left and at the top of the figure, respectively. Figure modified from Nelson-Sathi et al. (2015).

Methanosarcinales and Halobacteriales (>600 proteins). Interestingly, Methanosarcinales and Halobacteriales also represent the archaeal groups with the largest genomes, which might be explained in part by the high amount of shared genes with several members of Bacteria (see Figure 6.1). The distribution of bacterial homologs is mostly similar in all the members of Halobacteriales, whereas in Methanosarcinales, the high amount of bacterial homologs occurs predominantly in members of the Methanosarcina genus (intense red and large pattern in Figure 6.1). These findings are in agreement with previous reports suggesting that some members of Methanosarcina (Deppenmeier et al. 2002; Garushyants et al. 2015) and Halobacteria (Nelson-Sathi et al. 2012) have a largest set of genes from bacteria as product very likely of multiple HGT events.

6.2 Distributions of specific archaeal genes acquired from Bacteria correspond to the 13 archaeal higher taxa.

From the 25,762 protein families, 16,983 families were identified as archaeal specific, i.e. families with no homolog detected among the 1,847 bacterial genomes. A total of 14,416 families (85%) are occurring only in members of single higher taxa (on group), while in the remaining 2,567 families

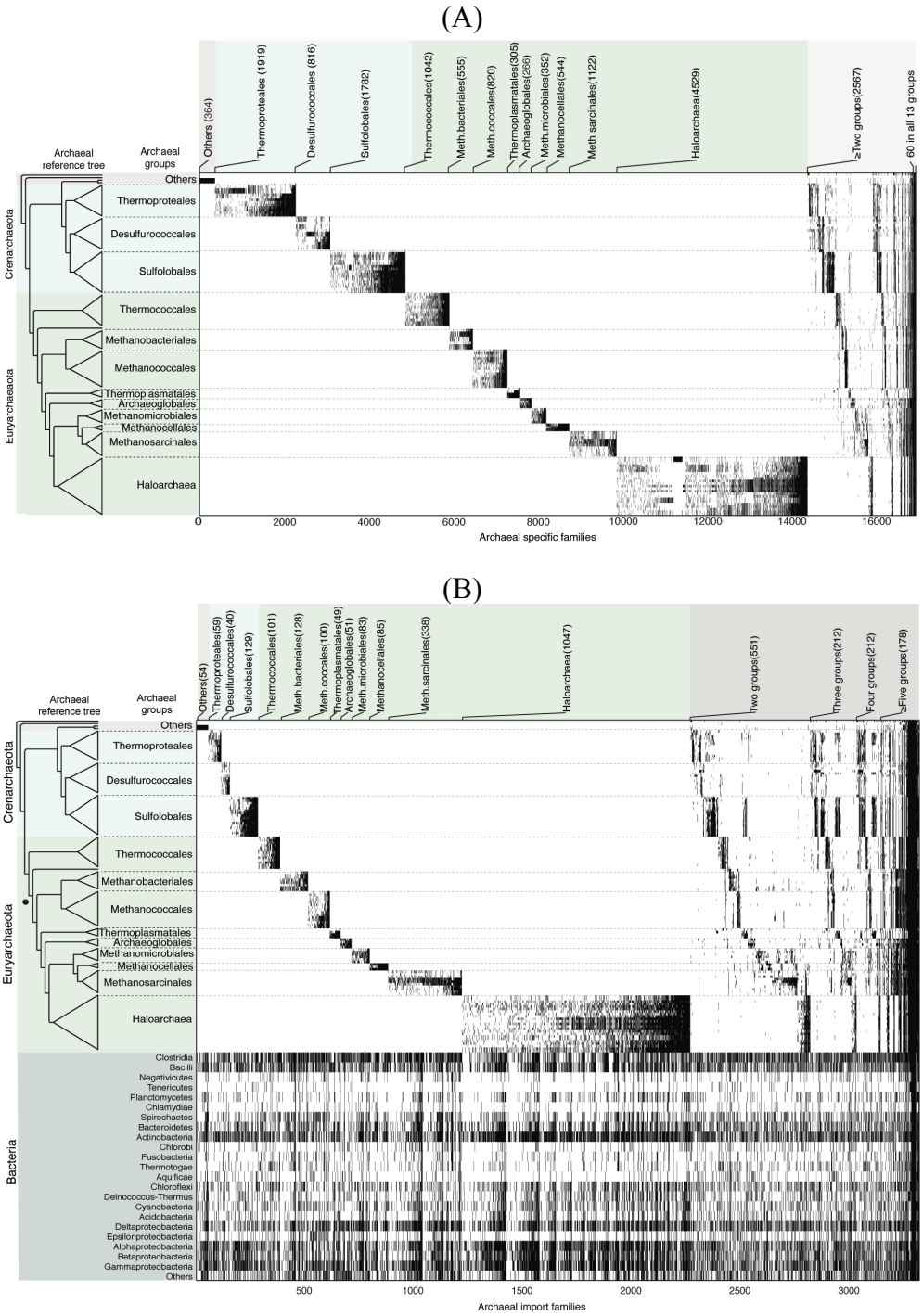


Figure 6.2. Distribution of genes across archaeal families. Archaeal-specific protein families (A) and archaeal protein families with bacterial homologs (B). For each cluster, ticks indicate presence (black) or absence (white) of the gene in the corresponding genome (rows, left axis). The number of clusters containing taxa specific on each group is indicated (upper legends). Reference tree for the archaeal genomes was obtained with 70 universal proteins is shown (left side), indicanting the names of the higher taxa groups in archaea. **Panel A:** Phylogenetic trees for 16,983 archaeal gene families were reconstructed (lower axis). Is important to note that 41,560 archaeal proteins did not have archaeal homologs, thus, they were classified as singletons and were excluded from further analyses. **Panel B:** 3,315 phylogenetic trees were reconstructed for each of the 8, 471 families with bacterial homologs containing ≥ 4 taxa. The ticks in the monophyletic trees (lower axis) indicate gene presence (black) or absence (white). The lower panel shows the occurrence of homologs among 23 bacterial groups. 2,264 genes are present in a single archaeal group (left portion), whereas only 56 genes are present in all 13 archaeal groups (fully black columns, right portion). Figures obtained from Nelson-Sathi et al. (2015).

Table 6.1. Functional annotation for archaeal genes according to gene family distribution and phylogeny*.

Function	COG category	Specific	M	NM	Exp	Imp
Information	Chromatin structure and dynamics	14	1	5	1	1
	Translation, ribosome biogenesis	263	84	50	9	27
	Replication, recombination and repair	375	126	185	17	69
	Transcription	524	124	113	10	81
Cellular	Defense mechanisms	48	62	116	4	45
	Cell cycle, division, chromosome partitioning	79	22	15	2	13
	Trafficking, secretion, vesicular transport	97	17	6	3	6
	Cell motility	146	40	29	8	33
	Cell wall/membrane/envelope biogenesis	197	143	203	10	91
	Protein turnover, chaperones	236	85	137	18	61
	Signal transduction mechanisms	308	120	129	16	101
Metabolism	Secondary metabolites	10	46	35	0	30
	Nucleotide transport and metabolism	44	53	105	7	41
	Lipid transport and metabolism	62	113	117	6	72
	Coenzyme transport and metabolism	168	143	219	11	97
	Inorganic ion transport and metabolism	232	176	265	16	123
	Carbohydrate transport and metabolism	118	205	227	14	139
	Energy production and conversion	334	254	403	25	175
	Amino acid transport and metabolism	177	278	440	26	208
No annotation	General function prediction only	949	434	560	49	297
	Function unknown	12602	789	715	139	554
Total		16983	3315	4074	391	2264

* *Categories for trees in the protein families are described in Methods section*

(15%) represent cases with ≥2 higher taxa (2,507), but no with the presence of all 134 species from Archaea. Only in 60 gene families, the 13 archaeal groups occur (see top headers at upper section A in Figure 6.2). The remaining 8,779 archaeal families have homologs with Bacteria representing different groups. Phylogenetic trees were reconstructed for those protein families with ≥ 4 proteins and then, a search of a single group of archaeal members (monophyletic group) in each tree was performed (see *import* definition in 'Detection of HGT events' section in Methods). In 4,397 trees, archaeal sequences were identified as monophyletic, while the remaining 4,074 trees were identified as non-monophyletic. The direction of the transfer was obtained by calculating the density distribution of the taxa by domain and for each family (*out of Africa* principle). Among the 4,074 monophyletic trees, 1,082 contain sequences from one bacterial genome or bacteria phylum and dominated by archaeal clades, thus these trees were classified as *exports* from Archaea to Bacteria. From the remaining 3,315 trees, 2,264 trees include genes represented by only one higher archaea taxa (center region at lower section B in Figure 6.2). Other 1,209 gene trees include genes represented by more than one higher archaeal taxa (right section at B in Figure 6.2). For example, 2264, 551, 212, 212, 178, and 56 genes occur specifically in one (65%), two (16%), three (6%), four (6%), more than five (5%), or in all archaeal (2%) groups, respectively.

Overall, the direction of the transfers shows a high asymmetry between Archaea and Bacteria, while 2,264 families were identified as *imports*, only 391 cases of exports were detected. Thus, the transfers from Bacteria to Archaea were fivefold more frequent than Archaea to Bacteria. Furthermore, in a sample-scaled set of equal number of bacterial and archaeal genomes, these transfers were 10.7-fold more frequent.

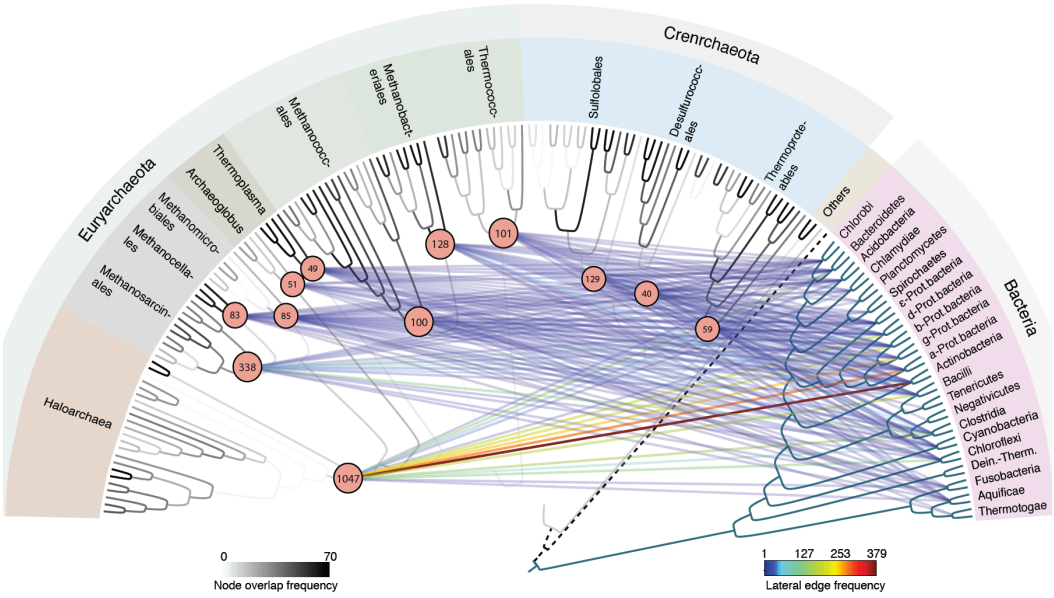


Figure 6.3. HGTs at the ancestral node of each archaeal clade and their bacterial clade donors. Archaeal genomes are organized according to the reference tree from Archaea, representing major phyla (Euryarchaea and Crenarchaeota), and 13 clades. The 2,264 transfers from Bacteria to Archaea are shown. The total number of imports is colored according to the lateral gene frequency from the bacterial donor (scaled bar on the right bottom level). Archaeal reference tree is presented by vertical edges, grey shading color from white (0) to dark grey (70) indicates how often the branch was recovered by the 70 gene analysed individually (scaled bar on the left bottom level).

6.3 Imported genes associated to metabolic functions in archaeal clades come from Firmicutes, Actinobacteria and Proteobacteria.

The transfers from Bacteria to Archaea are mainly related to "metabolic functions", where *amino acid import and metabolism* represents 208 gene families, *energy production and conversion* have 175 gene families, *inorganic ion transport and metabolism* have 123 gene families, and *carbohydrate transport and metabolism* represent 139 gene families. Also, "cellular functions" represent a high amount of imports, such as those involved in cell structure, e.g., *cell wall or envelope biogenesis* (91 genes), *chaperones* (61 genes), and *signal transduction mechanisms* (101 genes). Imports identified with "informational functions" represent in multiple COG categories, including *replication*, *recombination*, *repair* (69 genes) and *transcription* (81 genes). For a complete description see Table 6.1 and Supplementary Figure FS2.

The main bacterial donors of the 2,264 imports belong to different taxa that are represented mainly by four large groups: Proteobacteria (all classes), Bacilli, Clostridia and Chloroflexi (see Figure 6.3). In a minor amount, the remaining bacterial donor groups are Chlorobi, Bacteroidetes, Acidobacteria, Chlamydiae, Planctomycetes, Spirochaetes, Terenicutes, Negativicutes, Cyanobacgteria, Deinococcus-Thermus, Aquificae and Thermotogae. In archaea, there are clades having a minor amount of imports (<100), for example Thermoproteales (59), Desulfurococcales (40) in the Crenarchaeota phylum, and Thermoplasmatales (49), Archaeoglobales (51), Methanobacteriales (85) and Methanocellalles (83) in the Euryarchaeota phylum. In contrast, archaeal groups with a high amount of imports (≤100) are Sulfolobales (129) from Crenarchaeota, and Thermococcales (101), Methanobacteriales (128), Methanococcales (100), Methanosarcinales (338) and Halobacteriales (1,047) from Euryarchaeota.

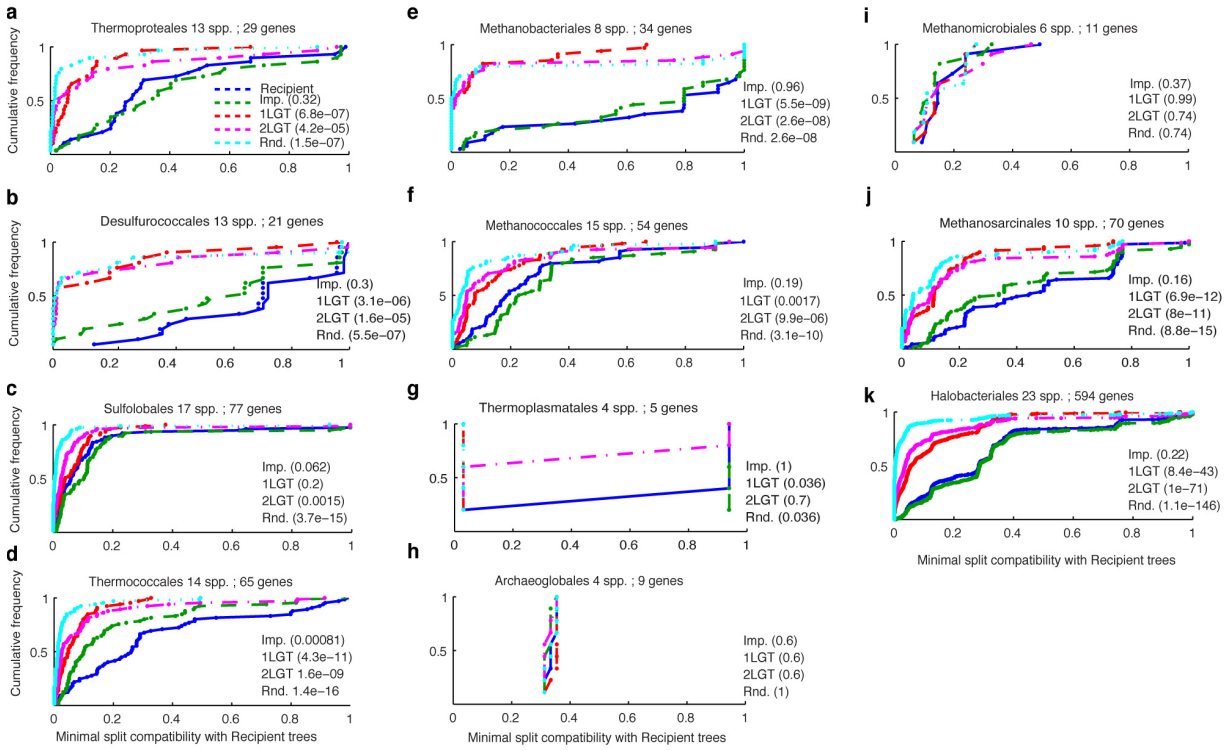


Figure 6.4. Comparison of sets of trees for single copy genes in 11 archaeal groups. Values are P values of the two-sided Kolmogorov-Smirnov (KS) two-sampled goodness-of-fit test in the comparison of the recipient (blue) data sets against the import (green) data set and three synthetic data set, one-LGT (red), two-LGT (pink) and random (cyan). The archaeal groups analized are Thermoproteales (**a**), Desulfurococcales (**b**), Sulfolobales(**c**), Thermococcales (**d**), Methanobacteriales (**e**), Methanobacteriales (**f**), Thermoplasmatales (**g**), Archaeoglobales (**h**), Methanococcales (**i**), Methanosarcinales (**j**), and Haloarchaea (**k**). The theoretical model used to compare tree sets and describing the *sampling size dataset*, *measurement of phylogenetic congruence*, and *the statistical test* are defined in the supplementary methods in Nelson-Sathi et al. (2015).

6.4 Comparative analysis of archaeal-specific and import gene trees shows that bacterial acquisitions correspond with the origins of archaeal clades.

In order to answer whether the origins of the archaeal groups coincide with the acquisitions of the imports, two single copy gene trees sets were used for this comparative analysis, a set of gene trees that do not show HGT (archaeal-specific) and, thus, they represent vertical inheritance, whereas and a second set of trees represent HGT (imports). For this purpose, a Kolmogorov-Smirnov (K-S) test adapted to non-identical leaf sets (split compatibility approach) was applied to these two collections of trees for each higher archaeal group, in which archaeal-specific genes (strong phylogenetic signal) against to imports (weak phylogenetic signal) were compared to test two main hypothesis. On the one hand, if both sets of trees (archaeal-specific and imports) evolved along the same phylogeny from a single origin, then both phylogenies should display the same phylogenetic signal (*null hypothesis*, H_0). On the other hand, both sets of trees (archaeal-specific and imports) not sharing a similar tree topology might suggest alternative phylogenetic scenarios, such as multiple horizontal gene transfers between lineages (*alternative hypothesis*, H_1). The pairwise comparison was carried in the 13 archaeal groups, and the significance (P, *p-value*) of each comparison in rejecting (*significant p-value*) or accepting (*non-significant p-value*) the null hypothesis for each archaeal clade is as follow: *Thermoproteales*

with 29 trees (P=0.32), *Desulfurococcales* (21 trees, P=0.3), *Sulfolobales* (77 trees, P=0.062), *Thermococcales* (65 trees, P=0.00081), *Methanobacteriales* (34 trees, P=0.96), *Methanococcales* (54 trees, P=0.19), *Thermoplasmatales* (5 trees, P=1), *Archaeoglobales* (9 trees, P=0.6), *Methanomicrobiales* (11 trees, P=0.37), *Methanosarcinales* (70 trees, P=0.16), and *Halobacteriales* (594 trees, P=0.22). From the 11 tested groups, only in 6 groups (*Thermoproteales*, *Desulfurococcales*, *Methanobacteriales*, *Methanococcales*, *Methanosarcinales*, and *Halobacteria*) the H_0 could not be rejected. That is, the comparison of set of trees strongly suggests that these imports were acquired at the ancestor of these clades. In addition, three synthetic set of trees were created to estimate a non-vertical inheritance by introducing a slight perturbation of the import set (one and two HGT events), while in a third set of trees, they were randomly sampled trees from the entire space of trees. In contrast, the slightest perturbation of the import set (one random prune and graft HGT event per tree) did reject the null hypothesis at $P<0.002$ in those six previous cases, very significant ($P < 10^{-42}$) for the Haloarchaea, with the largest sample of trees in the entire analysis. A graphical representation of these comparisons in each archaeal clade for both distributions is shown in the Figure 6.4.

6.5. Discussion of the chapter: HGT in the origins of archaeal clades.

As largely discussed (Achtman and Wagner 2008; Boucher et al. 2003; Dagan 2011; Doolittle and Papke 2006; Puigbo et al. 2010; Segerman 2012), multiple factors can contribute to the evolution of species. And yet, horizontal transfer is considered as one of the main mechanism of genomic variation directly related to the formation of species. In this research, the origin of archaeal clades was studied in the context of their bacterial gene acquisitions via HGT events.

To answer this question our study used three key important analysis to identify HGTs between bacteria and archaea: 1) the type or direction of transfers (either imports or exports), 2) a taxonomic classification of the recipient and donor groups of the transfers, and 3) the functions of the transferred genes (Nelson-Sathi et al. 2015). Based on this, our work describes the detection of 8,471 of HGT events between archaea and bacteria. From this total, only 4,397 HGTs correspond to very well defined cases were the direction of the transfers were identified as imports (2,264 trees) and exports (390 trees). However, in 4,074 trees the direction of the transfer was not identified. Here, we found that Euryarchaeota harbor more bacterial homologs than Crenarchaeota and, remarkably, Methanosarcinales and Haloarchaea represent the clades with the highest numbers of bacterial acquisitions, according to a previous report (Nelson-Sathi et al., 2012). Furthermore, we show that the amount of the HGT events in the two prokaryotic domains is highly asymmetric, e.g. transfers from bacteria to archaea are greater than fivefold more frequent than archaea to bacteria. Finally, we demonstrated that among all HGTs detected, bacterial acquisitions corresponds to the origin of each archaeal clade.

Overall, our findings support the theory of rapid expansion and slow reduction of genome evolution suggested by Wolf and Koonin (2013), through, for instance, subsequent gene loss by specific lineages, as was shown in Figures 6.2 and 6.3. It is important to to underline that is not clear whether that bacterial acquisition at the origin of the major archaeal caldes are the outcome of a major and unique HGT event or from multiple independent HGT events. In both cases, some well known mechanisms could be responsible of such transfers, such as independent HGT events (Lake and Rivera 2004; Williams et al. 2012), via unique combinations in founder lineage pangenomes (Fraser et al. 2008; Retchless and Lawrence 2007), or via mass transfers involving symbiotic associations,

similar to the origin of eukaryotes (McInerney et al. 2014; Williams et al. 2013). For lineages in which the origin of bacterial genes and the origin of the higher archaeal taxon are indistinguishable, the latter two mechanisms seem more probable.

7

HGT EVENTS AND GENE LOSSES HAVE SHAPED METHANOGENIC AND AQUATIC TRANSITIONS IN METHANOSARCINALES

Methanosarcinales species belong to one of the most diverse archaeal groups in terms of metabolism, physiology, and the ability to cope with different environments. In previous studies, multiple genes from different biological processes have been detected as a product of HGT events in few members of the Methanosarcina genus (Deppenmeier 2002; Garushyants et al. 2015; Nelson-Sathi et al. 2012; Youngblut et al. 2015). It reminds to be seen, however, to what extent HGT has shaped the evolution of the most distinctive biological processes in the Methanosarcinales clade, such as nitrogen fixation, methanogenesis, and halotolerance. Thus, a new analysis to identify and characterize HGT events across the most recent number of sequenced Methanosarcinales members (database version 2012) was accomplished and it is described in this chapter.

7.1 Accounting for HGTs in Methanosarcinales

Table 7.1 shows the distribution of the 3,128 protein gene families, from the 25,762 families described in Chapter 6, that have at least one gene member identified as horizontally transferred between Bacteria and Archaea. The Methanosarcinales genomes included in this analysis are represented by 6 different genera: *Methanosaeta* (3 species), *Methanosarcina* (3 species), *Methanococcoides* (1 species), *Methanosalsum* (1 species), *Methanohalobium* (1 species) and *Methanohalophilus* (1 species) (see Figure 7.1). The results obtained with this former data set of 10 Methanosarcinales species are going to be fully described next. As an updated and robust overview of the present results, a complementary analysis with additional 24 recent sequenced Methanosarcinales species (v.2017) is described at the end of this chapter.

According to the definitions described in the Methods section, a set of 1,045 protein gene families out of the 3,128 was identified as a product of HGT events transferred from Bacteria to Archaea. All these HGT transfers were designed as imports, since they form a monophyletic a group (see Table 5.2 in Methods). From the 1,045 imports, a total of 339 HGT events (32%) were further classified as *imports-Ms* given that these imports were transferred exclusively from Bacteria to the Methanosarcinales group. The remaining 706 imports (68%) were classified as *imports-Archaea,* since these HGT events were transferred from Bacteria to several archaeal species, including Methanosarcinales. A second set of 313 protein gene families out of the 3,128 HGT events were

identified as *exports*, *i.e.*, HGT transfers that most likely occurred from Archaea to Bacteria, although the specific archaeal donor and bacterial recipient were not resolved. Yet, it was possible to identify 152 exports as derived exclusively from Methanosarcinales (*exports-Ms*), whereas 161 exports are found to be derived from any of the other several archaeal clades involved, analyzed Methanosarcinales (*exports-Archaea*). Finally, a third set of 1,770 protein gene families from the 3,128 HGT events was identified as *non-monophyletic* cases, given that archaeal species represented in the phylogenetic tree (including Methanosarcinales) do not form a single group, but instead, they are interleaving with bacterial branches. Since these non-monophyletic transfers cannot be identified as imports or exports, they are considered as '*undefined transfers'* (see definitions on Table 5.2 and Figure 5.1 in Methods). From the 1,770 'undefined transfers', only 75 were exclusively found in Methanosarcinales, while the rest is shared between Methanosarcinales and other archaeal species (see Supplementary Figure SF3).

Since the main interest of this work is understand the evolutionary and ecological impact of the horizontal gene transfers occurred towards the Methanosarcinales species exclusively, a deep analysis of the 'exports' and 'undefined transfers' categories is beyond the scope of the present study. Furthermore, the HGT events defined as imports represent the most technically clear cases in phylogenetic trees analyzed, according to three main definitions: 1) the archaeal clades form a monophyletic group, 2) the direction of the transfer is detected, and 3) the most probable bacterial donors are identified. Thus, this chapter is mainly focused in the deep characterization of the 'imports' category in Methanosarcinales, as described next.

7.2 Differential distribution of 1,045 imports across the Methanosarcinales species.

In order to differentiate the distribution of imports among these genera, the species were divided in seven groups based in their taxonomy and location in the reference tree. As shown in Figure 7.1, all sequenced species from the *Methanosarcina* genera are depicted as 'Methanosarcina' (Msa), whereas all other sequenced species from the *Methanosarcinaceae* family that do not belong to the *Methanosarcina* genera are depicted as 'Others' (Others). Likewise, all sequenced species from the *Methanosaeta* genera are depicted as 'Methanosaeta' (Mta). In order to facilitate a comparative analysis, further *ad hoc* groups are thus depicted as the combinations of the previous ones, for instance: Msa-Others, Mta-Others and Mta-Msa.

Table 7.1 show the gene distribution of all HGTs classified as imports (1,045), exports (313), and as undefined transfers (1,770) across the archaeal species, according to the two major classes: the 339 imports located within the Methanosarcinales taxa exclusively (imports-Ms), and the 706 imports shared between Methanosarcinales and other archaeal clades (imports-Archaea). First, there is a differential distribution in the number of imports, either the total 1,045 or within each import class, across the archaeal species analyzed. For instance, around 17% (180) of the imports are observed in the *Methanosaeta* group, 35% (371) in Methanosarcina and 11% (111) in the group 'Others'. Similar proportions are also observed within each import class. In the 'imports-Ms' class, for instance, 46% of the HGT imports are observed in the Methanosarcina group, 11% in Methanosaeta, and 7% in the group 'Others'. In the 'imports-Archaea' group, 30% of the imports are present in Methanosarcina, 20% in Methanosaeta, and 12% in species of the group 'Others'. Around 16% (167) of the remaining HGT imports, in any of the previous categories (total imports, 'imports-Ms' or 'imports-Archaea'), are located in the Methanosarcinaceae family ('Msa-Othrs').

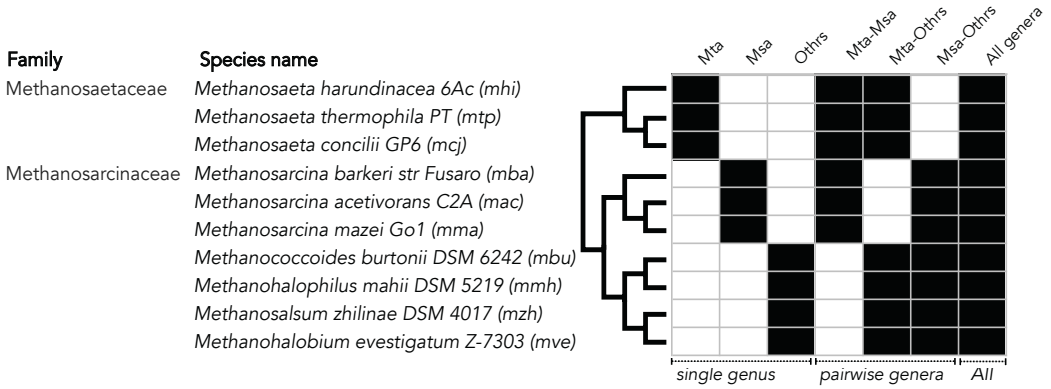


Figure 7.1. Composition of groups in the taxonomic order of Methanosarcinales used to differentiate the distribution of imports among species and genera. *Left side*: Methanosarcinales species are divided by taxonomic families and sorted based on the reference tree previously reconstructed (see Figure 6.1). The corresponding three-letters name from KEGG for each species is shown within brackets. *Center*: The phylogenetic relationships among Methanosarcinales species. *Right side*: Seven groups were created (columns), and the presence of each species is shown in black (filled box), while the absence of the species in the groups is represented as empty boxes. *At the bottom*: The tree first Methanosarcinales groups represent a "single genus", the next three groups are containing more than two genera ("pairwise genera"), while the last one (All) contains at least one member of each genera. The groups were defined as follow: ***Methanosaeta* (Mta)**: *M. harundinacea* (mhi), *M. thermophile* (mtp), and *M. concilii* (mcj); ***Methanosarcina* (Msa)**: *M. barkeri* (*Fusaro*) (mba), *M. acetivorans* (mac), and *M. mazei* (mma); ***Others group* (Othr)**: *M. burtonii* (mbu); *M. mahii* (mmh); *M. zhilinae* (mzh); *M. evestigatum* (mve). The first two groups (Mta and Msa) represent a single genus, while the third (Othrs) is a set of genera clustered together by a branch. For single genus groups, at least one protein of a genus must be present, whereas groups with ≥2 genera should contain, at least, 1 protein according to the corresponding genus.

TABLE 7.1. Distribution of imports in Methanosarcinales according to their specific groups and type of transfer: *imports-Ms* or *imports-Archaea*.

Class	Group	Species	Mta	Msa	Othrs	Mta-Msa	Mta-Othrs	Msa-Othrs	All	Total transfers
Imports			**[185]**	**[366]**	**[111]**	**[49]**	**[10]**	**[167]**	**[157]**	**[1045]**
	Methanosaeta	*M. harundinacea 6Ac*	104	0	0	22	4	0	139	269
		M. thermophila PT	70	0	0	15	5	0	135	225
		M. concilii GP6	115	0	0	37	9	0	152	313
	Methanosarcina	*M. barkeri str. Fusaro*	0	259	0	30	0	122	155	566
		M. acetivorans C2A	0	298	0	32	0	154	148	632
		M. mazei Go1	0	197	0	22	0	135	154	508
	Others	*M. burtonii DSM 6242*	0	0	42	0	4	118	145	309
		M. mahii DSM 5219	0	0	47	0	3	108	142	300
		M. zhilinae DSM 4017	0	0	49	0	5	90	137	281
		M. evestigatum Z-7303	0	0	40	0	4	101	143	288
Exports			**[57]**	**[99]**	**[23]**	**[13]**	**[10]**	**[49]**	**[62]**	**[313]**
	Methanosaeta	*M. harundinacea 6Ac*	41	0	0	8	8	0	60	117
		M. thermophila PT	26	0	0	5	6	0	54	91
		M. concilii GP6	37	0	0	10	6	0	60	113
	Methanosarcina	*M. barkeri str. Fusaro*	0	75	0	9	0	38	62	184
		M. acetivorans C2A	0	85	0	10	0	43	62	200
		M. mazei Go1	0	54	0	6	0	43	62	165
	Others	*M. burtonii DSM 6242*	0	0	8	0	3	35	57	103
		M. mahii DSM 5219	0	0	11	0	4	29	58	102
		M. zhilinae DSM 4017	0	0	11	0	2	24	55	92
		M. evestigatum Z-7303	0	0	11	0	5	33	59	108
Undefined			**[203]**	**[279]**	**[146]**	**[163]**	**[77]**	**[279]**	**[623]**	**[1770]**
	Methanosaeta	*M. harundinacea 6Ac*	139	0	0	110	52	0	588	889
		M. thermophila PT	78	0	0	65	32	0	510	685
		M. concilii GP6	158	0	0	130	53	0	627	968
	Methanosarcina	*M. barkeri str. Fusaro*	0	174	0	144	0	215	662	1195
		M. acetivorans C2A	0	227	0	140	0	271	695	1333
		M. mazei Go1	0	157	0	113	0	230	654	1154
	Others	*M. burtonii DSM 6242*	0	0	61	0	40	176	554	831
		M. mahii DSM 5219	0	0	50	0	32	156	550	788
		M. zhilinae DSM 4017	0	0	59	0	23	161	526	769
		M. evestigatum Z-7303	0	0	61	0	34	154	532	781

Header titles: **Mta**, Methanosaeta; **Msa**, Methanosarcina; **Othrs**, members of Methanosarcinaceae family; **Mta-Msa**; members of Methanosaeta and Methanosarcina; **Mta-Othrs**, members of Methanosaeta and from Othrs; **Msa-Othrs**, members of Methanosarcina and Othrs; **All**, members of all groups. The total number of gene families (square brackets), and the total nnumber of protein contained in the families are show per column. **Total transfers** are shown as the summary of all gene families in each type of transfer.

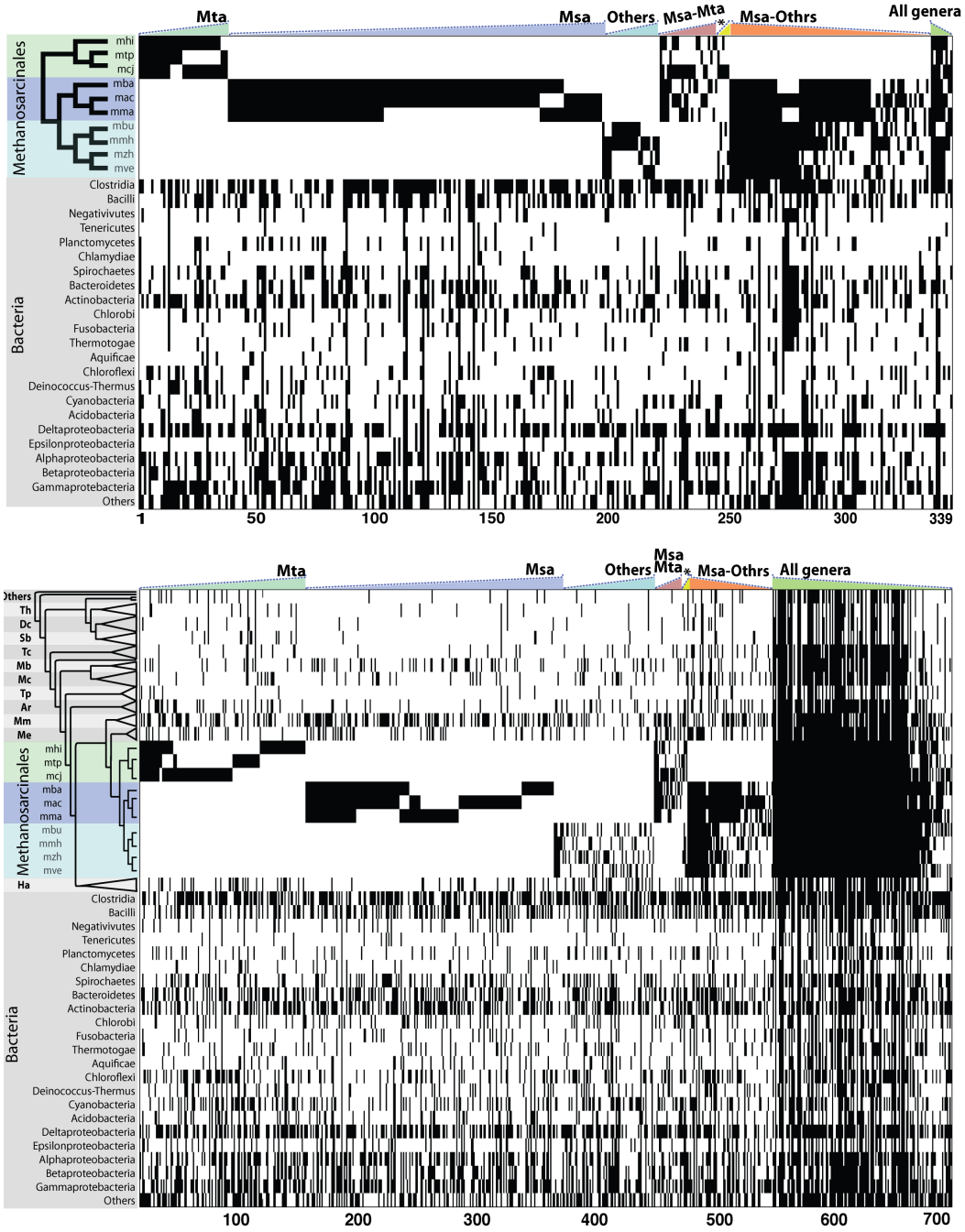


Figure 7.2. Gene distributions of 1,045 import in across all Methanosarcinales members. Imports are distributed by each Methanosarcinales subgroup (colored accordint to the top of each presence-absence figure). The asterisk (*) represents the Mta-Othrs subgroup in both upper and lower figures. **Upper figure:** distribution of the 339 *specific imports* across Methanosarcinales. **Lower figure:** distribution of the 706 *shared imports* across archaea. Archaeal genomes are listed based on the reference tree (see Figure 5.1). Archaeal genomes belonging to the same clade were collapsed to a single branch; only Methanosarcinales branches were kept expanded (colored boxes). ***Methanosaeta*** **(Mta)**: mhi, *M. harundinacea*; mtp, *M. thermophile*; mcj, *M. concilii*; mba; ***Methanosarcina*** **(Msa)**: *M. barkeri* (*Fusaro*); mac, *M. acetivorans*; mma, *M. mazei*; ***Others subgroup*** **(Othr)**: mbu, *M. burtonii*; mmh, *M. mahii*; mzh, M. zhilinae; mve, *M. evestigatum*. Archaeal clades are organized in 13 groups: Others; Th, *Thermoproteales*; Dc, *Desulfurococcales*; Sb, *Sulfolobales*; Tc, *Thermococcales*; Mb, *Methanobacteriales*; Mc, *Methanococcales*; Tp, *Thermoplasmatales*; Ar, *Archaeoglobales*; Mm, *Methanomicrobiales*; Me, *Methanocellales*, Ha, Haloarchaea. 23 Bacterial groups are displayed below to the Methanosarcinales tree. Ticks indicate the presence (black) or absence (white) of the gene in the corresponding either archaeal and bacterial genome. Bacterial groups are shown at the bottom of each figure.

Notably, the distribution of imports within the 'imports-Ms' class in Table 7.1 and in Figure 7.2 A shows that the most frequent recipients of HGT events from bacteria are the three species belonging to Methanosarcina: 220 in *Methanosarcina barkeri* (6.1% of the proteome), 246 in *Methanosarcina acetivorans* (5.4% of the proteome), and 179 *Methanosarcina mazei* (5.3% of the proteome). In terms of the archaeal proteome content, the genomes of the *Methanosaeta* genus harbor 13% (258 ± 24 genes from 258 protein gene families) of bacterial genes in average, whereas the genomes in *Methanosarcina* genus contain 34% (539 ± 30 genes from 539 protein gene families) of bacterial genes on average. Proteomes of the species grouped in the 'Others' class (*Methanococcoides*, *Methanohalophilus*, *Methanosalsum* and *Methanohalobium*) harbor around the 14% (262 ± 4 genes) of bacterial genes on average. These estimations are in well agreement with those studies based on bacterial homologs from both BLAST (Deppenmeier 2002) and phylogenetic approaches (Garushyants et al. 2015; Nelson-Sathi et al. 2012).

The distribution of imports within the 'imports-Archaea' class in Figure 7.2 B shows that the imported genes among the archaeal taxa occur more frequently in clades from Euryarchaeota than in those from Crenarchaeota. For instance, the main Euryarchaeota clades sharing imports with all species in Methanosarcinales are *Methanobacteriales*, *Methanomicrobiales* and *Methanocellales*. In particular, imports found in *Methanosaeta* and members of 'Others' are also found in the Euryarchaeota clades *Archaeoglobales* and *Methanococcales*. From Crenarchaeota, *Thermococcales* is the unique genus found to share imports with some Methanosarcinales species.

7.3 Proteobacteria and Firmicutes are the principal donors of the imported genes in Methanosarcinales.

In order to identify particular bacterial donors for a given HGT import, the 23 major bacterial taxa included in this study were subdivided into 44 minor taxonomic groups (see lists of taxa in Figure 7.3). Since a bacterial sister branch of the imported gene can harbor more than one closely and not-closely related bacterial species, three categories were created to define the potential bacterial donors. First, a 'single clade donor' was identified when the sister group is represented by species from one single clade represented by one of the 44 bacterial taxa described here, e.g., *α-proteobacteria*. A total of 518 HGT imports (50%) were assigned to a single clade donor. Second, a 'putative clade donors' was assigned when the sister group is represented by two or more closely related clades from the 44 bacterial taxa, e.g., *α-proteobacteria* and *β-proteobacteria*. A total of 52 HGT imports (5%) were assigned to a 'putative clade donors'. Third, 'undefined bacterial donors' were assigned when the sister branch is represented by two or more not closely related bacterial clades from the 44 taxa, such as Firmicutes and Proteobacteria. The remaining 457 HGT imports (45%) have undefined bacterial donors, which are not shown.

From the total 570 HGT imports for which was possible to identify a single or putative clade donors, the most frequent bacterial donors of Methanosarcinales include species from *Proteobacteria* (28%), *Clostridia* (24%), *Actinobacteria* (7%), *Spirochaetia* (4%), *Bacilli* (3%), *Planctomycetia* (3%), and *Thermotogae* (2%). Other bacterial donors of Methanosarcinales (albeit with limited occurrence) include species from *Chlorobi*, *Cyanobacteria*, *Ignavibacteria*, *Acidobacteria*, *Deinococcus-Thermus*, *Nitrospirae*, *Planctomycetes*, *Spirochaetes*, *Verrucomicrobia,* among others (see Figure 7.3 and Supplementary Table ST1). We further ask whether there are taxonomic similarities among the 570

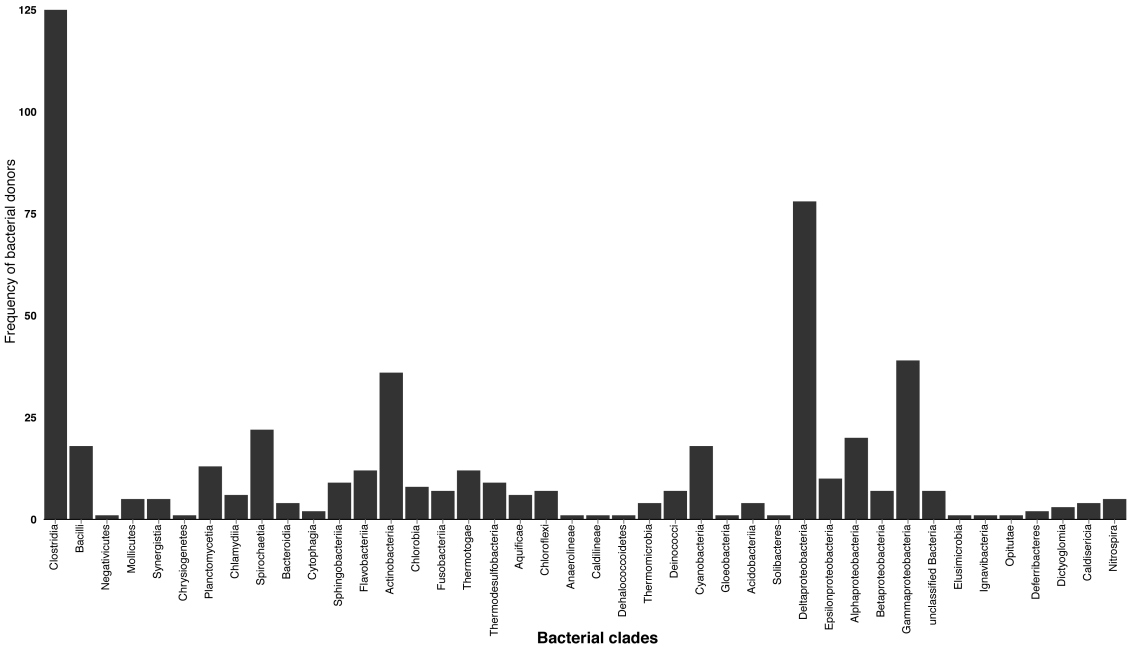


Figure 7.3. Bacterial donors of the imported genes in Methanosarcinales. Each bacterial donor clade was identified as a sister group (branch) of an archaeal monophyletic group in each phylogeny, either specific or shared imports. Bacterial sister group (donor) can be composed by ≥1 bacterial member of a single clade, such cases were defined as *single bacterial donor*. However, in several cases, a sister group cannot be well defined because is represented by ≥2 bacterial clades. In cases where the clades belong to the same higher taxonomic *phylum*, the putative donor group is defined by the phylum; these cases are named as *putative bacterial donors*. For example, a donor is assigned to Firmicutes phylum when two or more taxonomic groups, such Bacilli and Clostridia, are present in the sister branch.

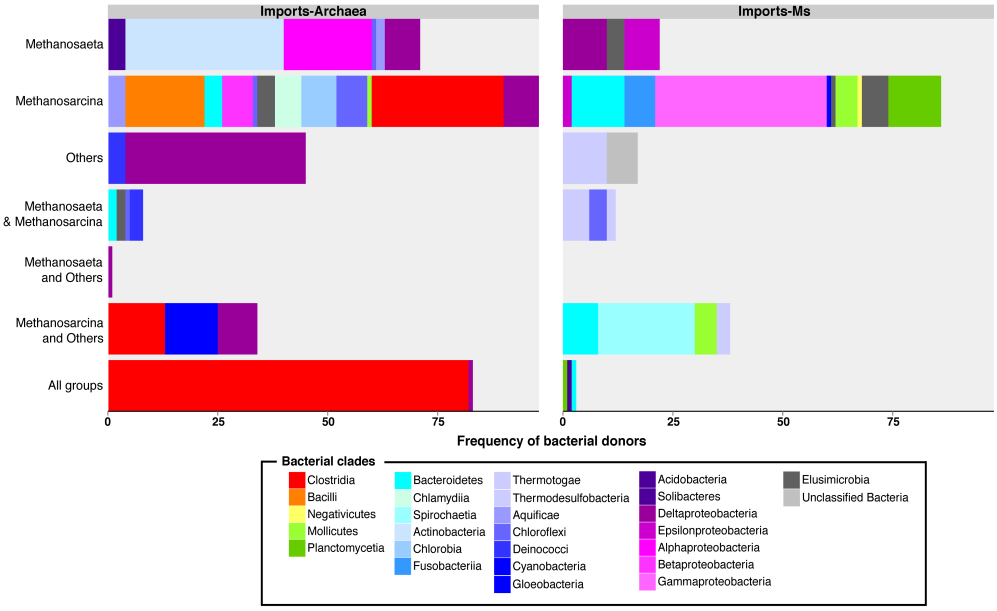


Figure 7.4. Identified bacterial donors of specific and shared imported genes in Methanosarcinales. Bacterial donors were separated by those found in the *Imports-Archaea* (left panel) and those belonging to the *Imports-Ms* (right panel). Methanosarcinales groups are listed from the top to the bottom at the left axis: **Methanosaeta (Mta), *Methanosarcina* (Msa), *Othrs*,** Mta-Msa, Mta-Othrs, Msa-Othrs, All groups (All genera). The 'x' axis depict the frequency of bacterial donors calculated by using the 570 detected sister clades on each Methanosarcinales branch. Bacterial groups (at bottom) are represented by colors. A complete list of bacterial donors is shown in the Supplementary Table ST1.

single and putative bacterial donors found in both 'imports-Ms' and in 'imports-Archaea'. As shown in Figure 7.4, it is possible to observe that *Methanosarcina* species not only have the highest number of both types of imports, but also posses a varied diversity of bacterial donors in comparison to other archaeal clades. Within the 'imports-Ms', all major clades from *Proteobacteria* represent the most frequent donors. Other exclusive and frequent bacterial donors to Methanosarcinales include species from *Thermodesulfurobacteria, Chloroflexi, Thermomicrobia*, *Bacteroidetes*, *Planctomycetia*, and some unclassified Bacteria. Within the 'imports-Archaea', some clades from *Clostridia, Bacilli,* and *Actinobacteria* appear to be exclusive bacterial donors of HGT events to Methanosarcinales. Also, species from *Bacteroidetes*, *Cyanobacteria*, *Deinococcus*, and some groups in Proteobacteria (such as δ-*proteobacteria* and γ-*proteobacteria*) appear to be common bacterial donors to several archaeal clades, including Methanosarcinales.

7.4 Energy processing, environmental interactions, and cellular processes are the most widely represented biological processes among the imported genes.

To assess the type of functions associated with the 3,128 protein gene families that that have at least one gene member identified as HGT between Bacteria and Archaea, a functional annotation was obtained from both COG and KEGG databases (see Methods). By using the most frequent functional annotation per gene family as criterion, the COG approach was employed to estimate a unique functional annotation for each protein family. In addition, the KEGG approach was used to identify the metabolic pathways involve in all single genes identified as HGT events. The total number of genes, functions and domain distribution for the 3,128 protein gene families is summarized in the Supplementary Figure SF5.

Although there is a clear over-representation of species from Bacteria (e.g. Firmicutes and Proteobacteria), 60% of the 3,128 HGT events (i.e., 'imports', 'exports' and 'undefined transfers') between Bacteria and Archaea domains are associated to *operational* functions. The most dominant category within the operational functions involves the *metabolic* ones (36%), followed by the *cellular* and *structural processes* (16%), and *informational* roles account for 12% of the total HGT events, including functions such as ribosomal proteins, translation, recombination and repair (see Table 7.2). Finally, around 35% of the HGTs are represented by poorly or not annotated genes, so that these transfers were assigned as '*unknown functions*'.

Figure 7.5 displays the functional distribution of all HGT imports ('imports-Ms' and in 'imports-Archaea') into 19 functional categories (as scanned from the KEGG database) and their corresponding proportion across the archaeal clades, including Methanosarcinales. On the one hand, this analysis shows that Methanosarcinales (21%) and Halobacteriales (19%) recapitulate most of the bacterial imports, while minor proportions of bacterial imports are captured by others archaeal clades such as *Sulfolobales* (9%), *Thermococcales* (7%), *Methanococcales* (11%) and *Methanomicrobiales* (7%). The smallest fraction of imports (≤5%) was found in *Thermoproteales*, *Desulfurococales*, *Methanobacteriales*, *Thermoplasmatales*, *Archaeoglobales* and *Methanocellales*. On the other hand, most of the HGT imports identified in Methanosarcinales are found to be involved in the amino acid metabolism, biosynthesis of secondary metabolites, xenobiotics biodegradation, signal transduction, transport and catabolism, and cell motility, replication and repair systems (see Figure 7.5). Overall, imports detected in the Methanosarcinales are associated to *energy processing* and *different types of*

Table 7.2. Functional annotation of protein families according to the phylogenetic trees classification for HGT detection.

Function	COG category	Import	Undefined	Export
Informational	Translation, ribosomal structure and biogenesis	52	32	14
	RNA processing and modification	0	1	0
	Transcription	27	48	13
	Replication, recombination and repair	76	102	16
	Chromatin structure and dynamics	3	2	0
	Cell cycle control, cell division, chromosome partitioning	7	8	1
Cellular	Defense mechanisms	29	55	1
	Signal transduction mechanisms	36	51	18
	Cell wall/membrane/envelope biogenesis	49	93	7
	Cell motility	10	15	3
	Intracellular trafficking, secretion, and vesicular transport	9	2	2
	Posttranslational modification, protein turnover, chaperones	26	71	11
Metabolic	Energy production and conversion	60	176	24
	Carbohydrate transport and metabolism	31	69	1
	Amino acid transport and metabolism	64	178	14
	Nucleotide transport and metabolism	9	53	2
	Coenzyme transport and metabolism	48	124	13
	Lipid transport and metabolism	9	25	4
	Inorganic ion transport and metabolism	59	129	14
	Secondary metabolites biosynthesis, transport and catabolism	15	12	1
No annotation	General function prediction only	127	249	42
	Function unknown	299	275	112
Total		1,045	1,770	313

* The numbers for imports represented both the types of *imports* (Imports-Ms and Imports-Archaea). The same counting approach was applied to *exports* and *undefined transfers* (non-monophyletic). In total, these HGTs are contained in 3,128 protein families. See supplementary Figures SF3 and SF4.

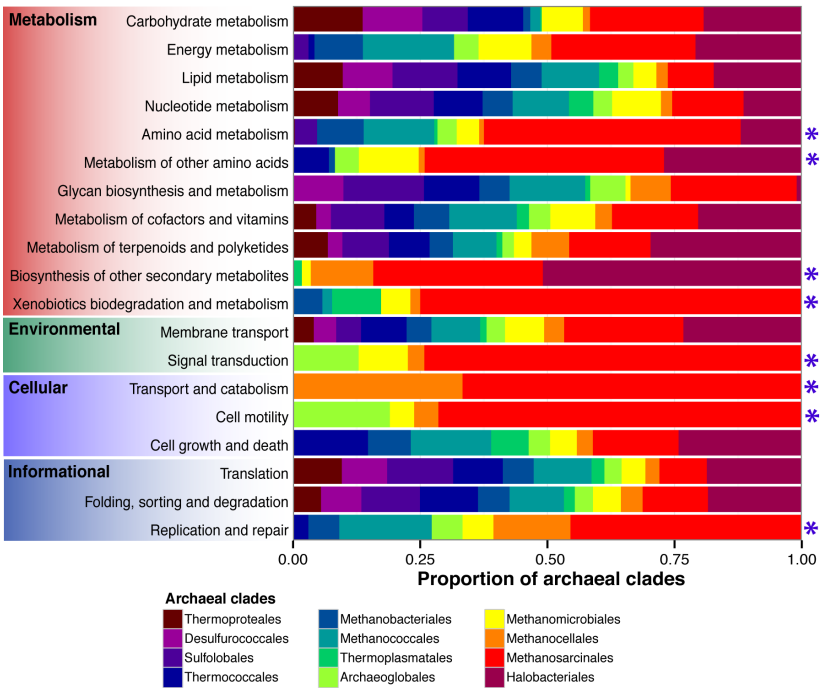


Figure 7.5. Distribution of imported genes across archaeal clades according to the functional categories of KEGG pathways. The imported genes represented by 1,045 families were classified in 19 functional categories of the KEGG pathways (axis at *left* side). The size of the horizontal bars represents the fraction of imports (*lower* axis) corresponding to any of the 12 specific archaeal clades (color code at bottom). Imports corresponding to the phyla *Korarchaeota*, *Nanoarchaeota* and *Thaumarchaeota* were not identified using the KEGG annotation.

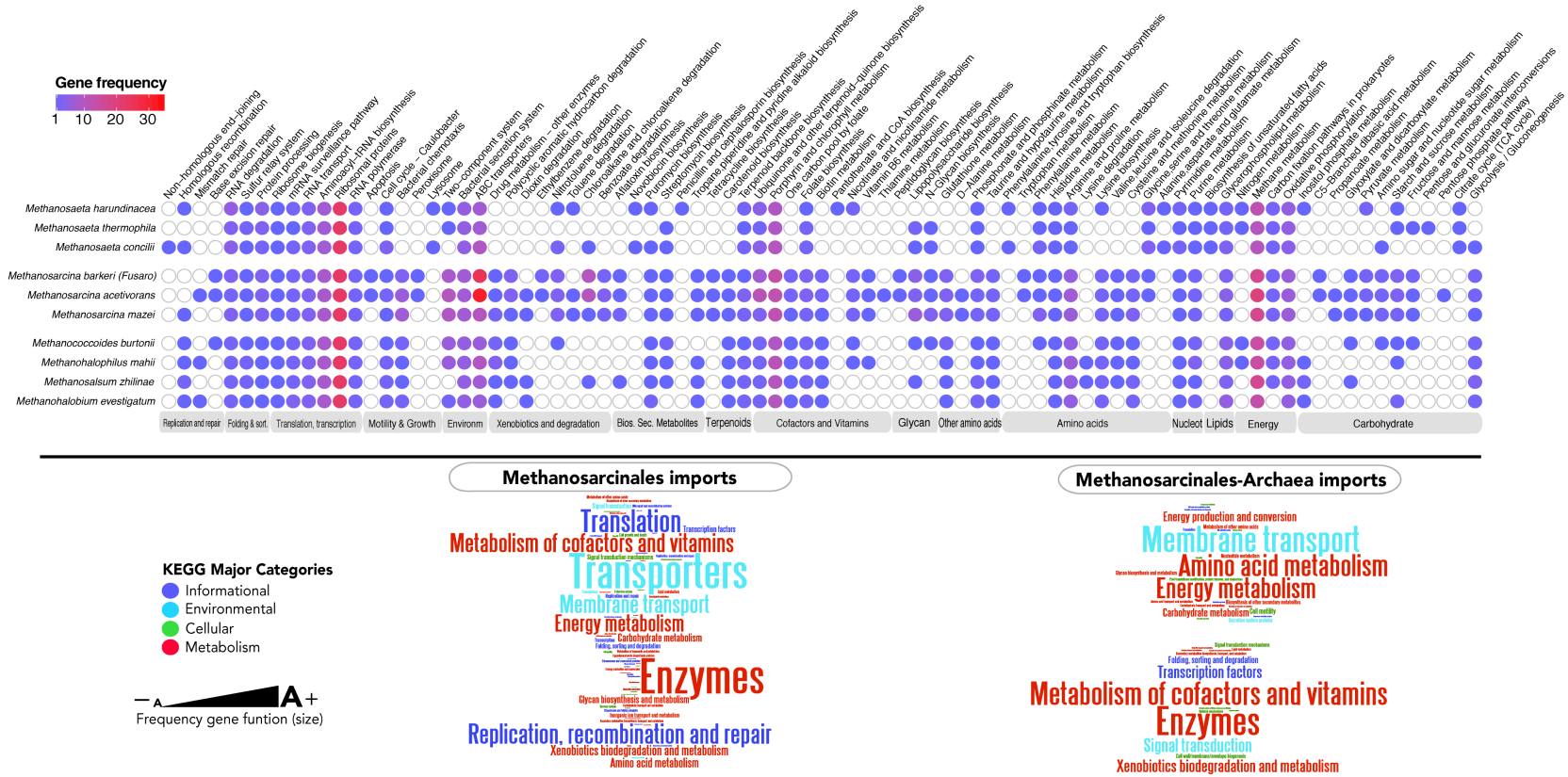


Figure 7.6. Metabolic pathways associated to the imported genes in Methanosarcinales members. Distribution and frequency of imported genes across all metabolic pathways found in Methanosarcinales species are shown at the top, while gene functions and its frequencies for either specific and shared imports are classified under different main classes at the bottom. **Upper panel:** Genes present across, and within categories are shown (filled circles); absences are represented as non-filled circles (grey outline circles). The frequency of genes per category was counted for each Methanosarcinales species (see colored-bar at the left-top). Biological functions present in all Methanosarcinales members (bold-black legend), and almost in all Methanosarcinales members (bold-turquoise legend) are shown. Metabolisms and other cellular processes are shown with the corresponding category (see grey labeled boxes at bottom). Biological processes are represented at the bottom. **Lower panel:** Functional KEGG classes for specific and shared imports (in colors), and their frequencies symbolized by a *Wordle* representation. Here, the word size relates a high or low frequency appearing within the annotation by increasing or decreasing its size, respectively. Major KEGG categories were used to divide genes associated to processes such as informational, environmental, and cellular, or to different metabolisms (colored labels at the left side).

metabolisms (30%), *multiple cellular processes* and *structural functions* (18%), *environmental interactions* (11%), and *informational functions* (9%).

As shown in Figure 7.6, differential functions are observed for the imported genes within the clades of Methanosarcinales. Exclusive functions identified in the Methanosaeta species are, for example, biosynthesis of novobiocin, pantothenate, CoA, and unsaturated fatty acids. Particular functions within the Methanosarcina group are ethylbenzene degradation, biosynthesis of peptidoglycan, pentose phosphate pathway, and diverse type of metabolisms, such as thiamine, D-alanine, tryptophan, and propanoate. Interestingly, those imports involved in lysine degradation are only found in the remaining Methanosarcinales groups. As also observed in Figure 7.6, multiple pathways and metabolisms are shared among several Methanosarcinales species and groups, such as ABC transporters, biosynthesis of quinone and folate, oxidative phosphorylation, and multiple metabolisms associated to amino acids and methane (within the *metabolic functions*), as well as the biosynthesis of aminoacyl-tRNA, RNA transport, and ribosomal proteins (within the *informational functions*). From the latest category, it is noteworthy to mention that 24 bacterial imports were identified as ribosomal proteins, and they are present in all Methanosarcinales species and in other multiple archaeal clades. From this set of ribosomal proteins, 13 are described as small (30S) ribosomal protein (rpsBCDEGHIJKLMNQS), 10 genes were identified as large (50S) ribosomal protein (rplABCEFLMNVW), and 1 gene is described as a putative RNA binding protein containing a KH domain. Other informational genes include transcriptional regulators (e.g. MarR, HxlR, ArsR, TrmB), helicase subunits, few integrases, CRISP-associated proteins and numerous transposases.

7.5 HGT at the hearth of versatile biological processes in Methanosarcinales: Electron Transport Chain and Nitrogen fixation, and Methanogenesis.

The findings described above provide evidence of numerous HGT events potentially involved in the metabolisms of methane, nitrogen, and energy in Methanosarcinales. To assess more accurately the impact of HGT in these metabolisms, a selected and literature-curated set of 188 *reference genes* related with these metabolisms was re-annotated by screening them for orthology, HGT, bacterial donor type and functional classification across all genomes of Methanosarcinales (see Methods and Tables 5.3-6). Noteworthy, this set of genes contains genes copies declared as, either expressed and having a functional role, or copies with potential role. Gene copies are found in the formylmethanofuran dehydrogenase in either tungsten (*Fwd*) or molybdenum (*Fmd*) forms, the $F_{420}H_2$ dehydrogenase complex (*Fpo*), the hydrogenases F_{420}-reducing hydrogenase (*Frh*) and F_{420}-non-reducing hydrogenase viologen (*Vht*), as well as other proteins involved in methanogenic parthways, such as the methylotrophic (methyltransferases) and aceticlastic (the AMP-forming acetyl-CoA synthetase (*ACS*), and the carbon monoxide dehydrogenase/acetyl-CoA decarbonylase complex (*CODH/ACS*)). From this set of reference genes, 93 genes (listed on Table 5.3) belong to membrane-bound proteins, energy production and nitrogen fixation (72 functions), 11 genes represent the set of genes (*operon*) related in to the biosynthesis of pyrrolysine (*pylSBCD*) and F_{420} (*cofC*, *cofD*, *cofE*, *cofG*, *cofH*), and 84 genes are directly involved in the methanogenesis pathways (50 functions) (listed on Table 5.4). From the 5,033 protein-gene families identified in Methanosarcinales, the results show that only 186 (out of the 188) reference genes are represented by 192 families. The genes not identified are the formate dehydrogenase (*fdhD*), the methenyltetramethanopterin hydrogenase (*Hmd*), and the subunits *fwdH* from the formylmethanofuran dehydrogenase. Likewise, genes playing a role in the methanogenesis were detected in 84 protein families, and only the genes *fwdH* and *hmd* were not found.

From the total set of 188 reference genes, 41 genes have 2 copies (20 in methanogenesis and 21 in membrane-bound proteins), particularly those related to methanogenesis: *cdhA*, *anfG*, *fwdA, fwdB, fwdC, fwdD, fwdE, fwdF, fwdG*, and *ACS*. Interestingly, some genes associated to the methylotrophic pathway have from 3 to 4 copies, such as *mtaABC*, *mtmBC*, *mtbB* and *mttC*. The maximum number of copies (5) was found for the subunit *mtrA*. Similarly, multiple genes copies related to the energy metabolism and nitrogen fixation were found in diverse protein complexes, such as A-ATPase (*ahaC, ahaE, ahaF, ahaI*), F_{420} H_2 dehydrogenase (*fpoA, fpoD, fpoF, fpoJ, fpoL*), nitrogenase (*nifDKENI*), the F_{420} non-reducing hydrogenase viologen (*vhtACG*), heterodisulfide reductase (*hdrABC*), and the antiporter Na^+/H^+ (*mrpC, mrpD, mrpF*). By contrast, some relevant proteins and membrane subunits are present in a *single copy* such as the *F-ATPase*, energy conserving hydrogenase (*Ech*), pyrrolysine operon (*Pyl*) and Rhodobacter nitrogen fixation complex (*Rnf*), and the subunits Fpo and *Mrp.*

From the total of 192 protein-gene families that have homologs of the reference genes, 144 families represent cases of HGT events, while the remaining 48 families represent orthologous genes within Archaea (i.e., archaeal specific). The 144 protein-gene families with HGTs were classified as imports (56), exports (23), and undefined (65) transfers. In the following sections, an explanatory analysis of these results is described for each metabolism and pathway.

7.5.1. Multiple subunits of energy membrane-bound complexes and protein transporters in Methanosarcinales are the outcome of HGT events.

Figure 7.7 shows the distribution of 84 homologous genes across the genomes of methanogenic species, including 9 membrane-bound and 2 cytoplasmic protein complexes, 4 single membrane proteins transporters, and 2 operons. It is possible to observe that some complexes are conserved between Methanosarcinales and other methanogens. Examples include formate dehydrogenase (*fdh*), heterodisulfide reductase (*hdrABC*), energy conserving hydrogenase (*ech*), A-ATPase (*aha*), and the nitrogenase (*nif*). On the other hand, other protein complexes seem to be exclusive of Methanosarcinales (see Supplementary Figure SF6). Examples include *vhtGAC,* Rnf (except for *RnfY*), the antiporter *Mrp* of Na^+/H^+ (except for *mrpC*), the F_{420}-non-reducing hydrogenase (*fpo*), the methylamine substrate uptake (*mtmP*, *mtbP*, and *mttP*), and the complete gene machinery to produce pyrrolysine (*pyl*).

Additionally, the direction of the transfer (i.e. import or export) was possible to be assigned for several complexes and protein-gene sets within Methanosarcinales. Most bacterial imports are located in the protein complexes *nif* nitrogenase, F-type ATPase (*atp*), and Rnf (Methanosarcina), as well as in membrane transporters of methylamine substrates, and pyrrolysine synthesis (Methanosarcina and the "Others" group). Most archaeal exports are located in several subunits from the A-type ATPase (aha) complex. Nevertheless, the identification of donors and recipients for most genes was no defined. After a manual inspection of 65 gene trees, only 6 HGT events were able to be re-defined as imports (*fdhC, hdrD, vhoA, anfG*, *mrpG,* and *fpoN*), and 5 HGT events were re-defined as exports (including the complexes A-ATPase and *ahaA, ahaB, ahaD, ahaI, ahaK*). Onlt 54 genes continued as undefined transfers, which suggest that multiple transfers of these genes might have occurred between archaea and bacteria. A manual inspection of the gene families also showed that some genes have paralogs either within the same family or in another protein family.

Table 7.3. Number of gene copy members on each membrane protein complex according to their gene family and type of gene category: *archaeal-specific* or *HGT* (continues next page).

Complex	*Location/substrate*	*Gene*	*Total*	*S*	*I*	*E*	*U*	*Gene family*
F-ATPase (atp)	membrane-bound	atpB	1		1			24440
		atpF	1		1			22248
		atpE	1		1			24588
		atpA	1		1			23933
		atpD	1		1			24322
		atpH	1		1			23308
		atpC	1		1			24137
		atpG	1		1			23955
		atpI	1		1			23900
A-ATPase (aha)	membrane-bound	ahaA	1				1	34
		ahaB	1				1	33
		ahaC	2		1	1		1055 | 1247
		ahaD	1				1	38
		ahaE	2	1			1	1063 | 1255
		ahaF	2	1			1	1075 | 1328
		ahaH	1	1				6547
		ahaI	2		1		1	1015 | 927
		ahaK	1			1		6661
Fpo	membrane-bound	fpoA	1		1			1500
		fpoB	1				1	416
		fpoC	1				1	994
		fpoD	2				2	254 | 924
		fpoE	1				1	5507
		fpoF	**2**	1			1	3703 | 8249
		fpoH	1				1	407
		fpoI	1				1	636
		fpoJ	2	1	1			7997 | 9925
		fpoK	1				1	1430
		fpoL	**2**				2	3430 | 681
		fpoM	1				1	760
		fpoN	1				1	1268
		fpoO	1	1				8642
Ech	membrane-bound	echA	1				1	5770
		echB	1				1	5847
		echC	1				1	601
		echD	1				1	7756
		echE	1				1	2493
		echF	1				1	5909
Hdr	cytoplasmic	hdrA	2	1			1	4766 | 942
		hdrB	3		1		2	3319 | 4645 | 718
		hdrC	4		2		2	1335 | 14127 | 3186 | 6741
	membrane-bound	hdrD	1				1	4064
		hdrE	1	1				6047
Rnf	membrane-bound	rnfA	1		1			11688
		rnfB	1		1			11630
		rnfC	1		1			11441
		rnfD	1		1			11513
		rnfE	1		1			11642
		rnfG	1		1			11400
		rnfX	1				1	1544
		rnfY	1	1				11454
Mrp	membrane-bound	mrpA	1		1			3702
		mrpB	1		1			3925
		mrpC	2		1		1	3867 | 804
		mrpD	2				2	2841 | 3207
		mrpE	1				1	1309
		mrpF	2		2			3866 | 3910
		mrpG	1				1	3034

Table 7.3 (continue from last page). Number of gene copy members on each membrane

protein complex according to their gene family and type of gene category: *archaeal-specific* or *HGT*.

Complex	*Location/substrate*	*Gene*	*Total*	*S*	*I*	*E*	*U*	*Gene family*
Vht/vho	membrane-bound	vhtA	2		1		1	14353 | 4497
		vhtC	2	2				15576 | 8721
		vhtD	1				1	2717
		vhtG	2			1	1	14827 | 1597
Fdh	membrane-bound	fdhA	2				2	303 | 9494
Fdh		fdhB	1		1			1755
Fdh		fdhC	1				1	612
Cofactor 420	Biosynthesis	cofC	1		1			652
(operon)		cofD	1		1			627
		cofE	2				2	2872 | 607
		cofG	1				1	625
		cofH	3		1	1	1	2085 | 540 | 9696
Pyl (operon)	pyrrolysine operon	pylB	1		1			8332
		pylC	1		1			9060
		pylD	1		1			9094
		pylS	1		1			9038
Permeases	Acetate	aceP	1				1	984
	Monomethylamine	MtmP	2	1	1			13473 | 9216
	Dimethylamine	MtbP	1		1			10235
	Trimethylamine	MttP	1		1			9118
Nif	Nitrogenase	nifD	2		1		1	17666 | 1817
		nifE	2			1	1	18200 | 2262
		NifH	2				2	1704 | 971
		nifI	2		1		1	1766 | 20777
		nifK	3		2		1	14184 | 17648 | 20810
		nifN	2		1		1	3107 | 7206
		anfG	2		2			17226 | 24470

Gene family category: *S, Specific; I, Import; E, Export; U, Undefined transfer*

Thus, genes copies present in different families might have been assigned also to different categories regarding the direction of the HGT events. So that, for example, some copies of the subunits *fpoJ*, *ahaE, ahaF* and *hdrA* (and other genes of methanogenesis, see Table 7.3) might be annotated as "imports", while copies can be annotated as 'exports' or 'undefined'.

Furthermore, protein sequence-based domains (from the PFAM database, see Methods) were first identified in the 188 protein-coding reference genes (see Figure 7.8), and also for every ortholog and xenolog counterpart to assess the robustness of the functional annotation. As observed in Figure 7.9, the majority of the orthologous and xenologous protein-genes within the methanogenic pathways across all methanogenic species have the same type and number of Pfam domains from the corresponding reference protein-gene. A minor and heterogeneous proportion of genes, such as *fwdF*/*fmdF*, *mtsA*, *mtsB* and *aha*, exhibit more Pfam domains (and even double in number) with respect to its corresponding reference gene. Another genes, also in minor proportions, have less Pfam domains than their corresponding reference protein genes, such as *fpoE*, *fpoF*, *hdrD*, and *MtbP*. Interestingly, proteins with higher numbers of domains occur in species members from Methanosaeta and in some *Methanosarcinaceae*. Overall, the conservation of these domains suggests that these genes might keep the catalytic function that has been reported for the reference model species.

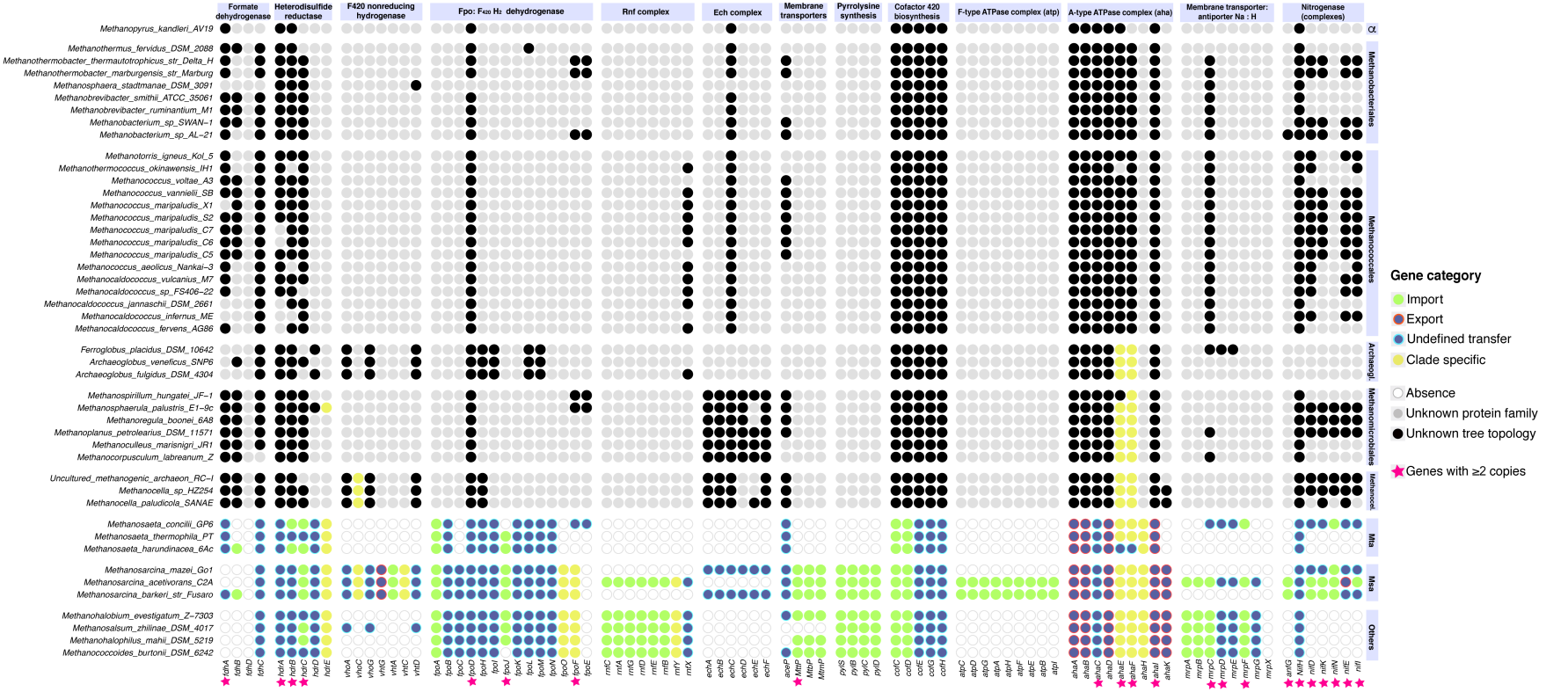


Figure 7.7. Membrane-bound complexes and transporters associated to methanogenesis and nitrogen fixation in Methanosarcinales. Twelve components are shown: 10 membrane-bound complexes, 4 single permeases, and the pyrrolysine operon. At the top (from left to right): Formate dehydrogenase, *fdh*; Heterodisulfide reductase, *hdr*; two F_{420} non-reducing hydrogenases, *vho* and *vht*; Rnf complex, *rnf*; Energy conservation hydrogenase, *ech*; acetate (*aceP*), tri- (mttP),di- (mtbP), and mono- (mtmP) methylamines uptake (substrates); Pyrrolysine operon, *pyl*; two ATPases (*atp* and *aha*); antiporter Na^+/H^+, *mrp*; and the Nitrogenase, *nif*. At center: Presence (filled circles) or absence (empty circles) of genes in the Methanosarcinales families is shown (see *Gene categories*). HGT events are described based on the tree topology (*Import*, *Export*, and *undefined transfer*), while its absence represents families without bacterial homologs (*clade specific*). Among the families, genes present in other methanogenic archaea with bacterial (black circles) or without bacterial homologs (yellow) are shown. Genes which were not found in the Methanosarcinales families, but its *presence* could be likely in other non-Methanosarcinales families, were considered as genes with *unkown protein family* (grey circles), whereas genes present in Methanosarcinales families but its *monophyly* was not tested, were designed as *undefined tree topology* (black circles). At bottom: Genes with ≥2 copies are shown (pink stars). The number of homologous proteins out of the Methanosarcinales families (i.e. 'unkown protein family') was estimated only for 48 genes, and is shown in the Supplementary Table ST2.

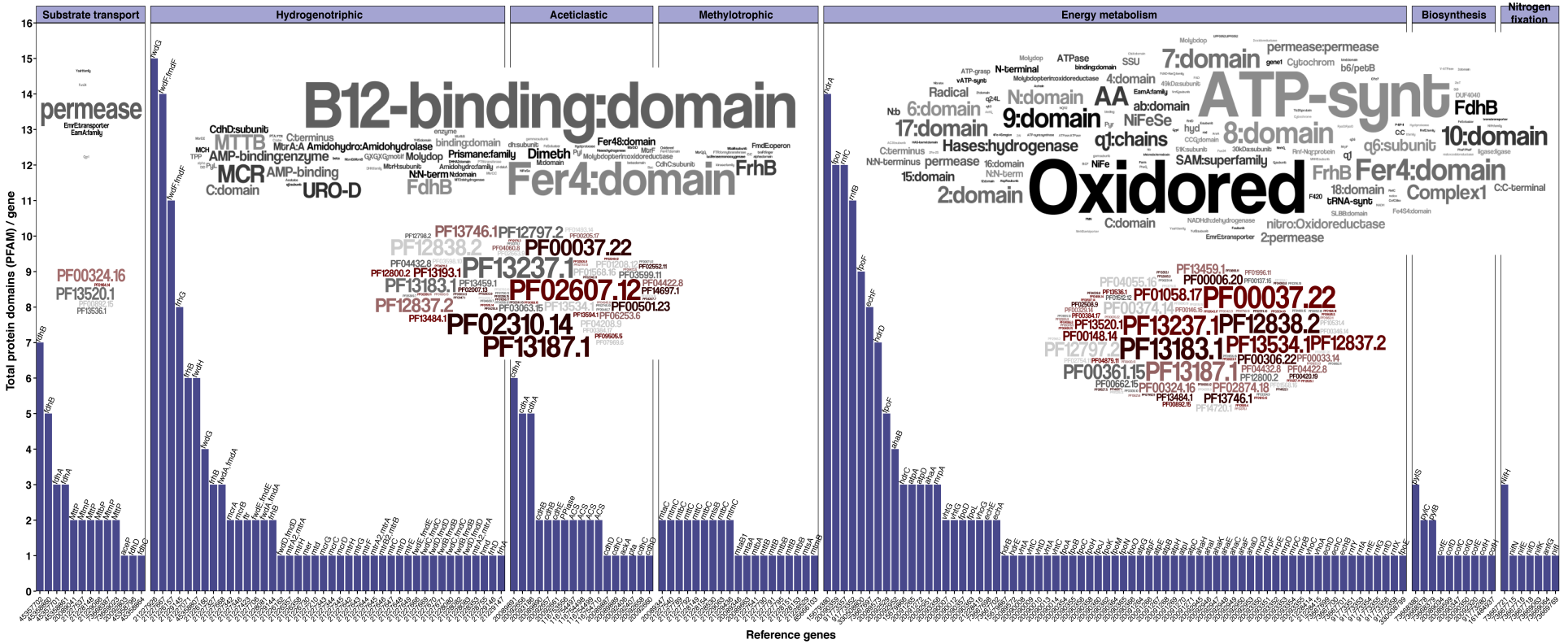


Figure 7.8. Protein domains identified in the reference genes involved in methane and energy metabolisms. The type and number of protein domains contained in 177 reference genes (lower axis) were identified using the Pfam database (see Methods). Protein domains were not identified only in 11 genes, such as *fpoJ* and *fpoO* (Fpo dehydrogenase), and the methanogenic genes *fwdE* (hydrogenotrophic), *mtmB*, *mtbB*, *mttB*, *mtmP*, *pylD* (methylotrophic) and *cdhE* (aceticlastic). The number of protein domains contained in the reference genes (*bars*) is shown ordered from maximum to minimum. Reference genes are divided in different functional roles (*legends* at the top): (from left to right) *methanogenic substrates* (methylated compounds, acetate or formate), *core enzymes of each methanogenic pathway* (hydrogenotrophic, aceticlastic, methylotrophic), *electron transfer chain respiration* through the membrane, *biosynthesis of cofactor*$_{420}$ and *pyrrolysine*, and *nitrogen fixation*. At the center, the type of protein domains and pfam accession numbers are symbolized by a Wordle representation (word size). Here, the word size relates a high or low frequency appearing within the annotation ('domain function' or 'pfam accessions') by increasing or decreasing its size, respectively.

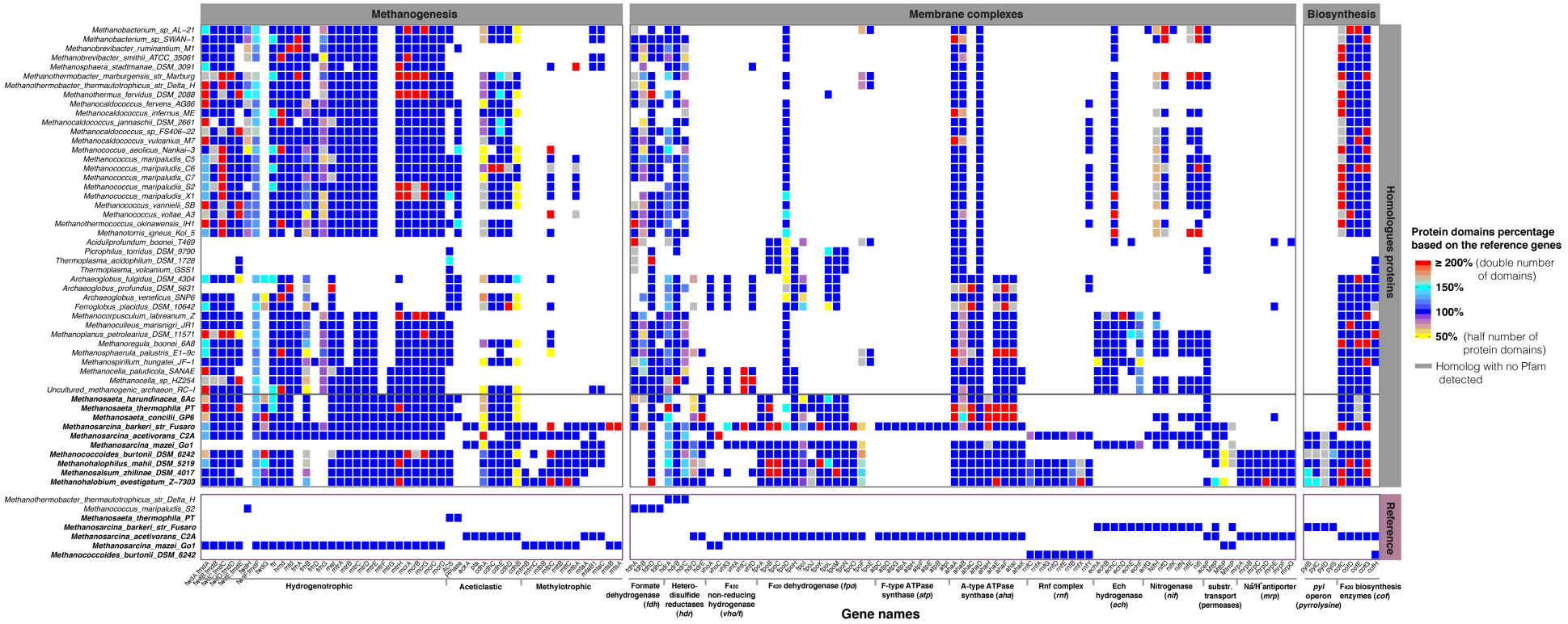


Figure 7.9 Identification of functional protein domains across genes involved with the methane and energy metabolisms in methanogens. A total set of 139 reference genes (see the lower panel *Reference*) involved with methanogenesis and energy production was analyzed in the context of their content of functional protein domains. A search of the type and content of protein domains in proteins involved with HGTs (see the upper panel *Homologues proteins*) was carried out using two protein domain databases (PFAM and SFAM, see methods). To define an operative or functional protein, i.e. a minimum number of domains in a functional protein (F_d), the set of reference genes was used as a frame of reference for such estimation (see the lower panel *Reference*). Among all reference genes, the most frequent number of domains contained in a protein was one, while others contain ≤15 domains, such as *fwdFG*, *hdrAC, hdrE*, *mrpA*, *rnfBC*, and *frhBG*. Becauset F_d is not uniform and can change in each reference gene, a normalization in percentage was used to illustrate the variability of F_d across multiple methanogens. Homologues genes containing the same number of domains (100% in blue), as well as those with more (above to 100%) or less (below to 100%) number of domains are shown. In homologues proteins, a F_d was calculated as follow: $F_d = \sum(D_h) \cdot (100)/\sum(D_r)$, where $\sum(D_r)$ is the total number of domains in a reference gene, $\sum(D_h)$ is total number of domains in a homologous gene, and 100 is the number to normalize in percentage.

7.5.2. Genes playing a role in the first steps of aceticlastic and methylotrophic methanogenesis are involved in HGT events in Methanosarcinales.

Figures 7.10, 7.11 and 7.12 display the distribution of The HGT events and other homologous counterparts identified for the reference genes on the aceticlastic, methylotrophic and hydrogenotrophic pathways across Methanosarcinales and other methanogens, respectively. Noteworthy, most of the homologous counterparts of the reference genes for the three methanogenic pathways conserve the same type and number (or even more) of Pfam protein domains (see Figure 7.9). Exceptions include the subunits *fwdH*, *frhB*, *cdhA*, and *cdhB*, which contain less Pfam domains than their corresponding reference genes.

As observed in Figure 7.10, the first reaction to be accomplished in the **aceticlastic methanogenesis** is the acetate activation to acetyl-CoA, which is carried out by either the *ackA/pta system* in Methanosarcina species, or by the *AMP-forming acetyl-CoA synthetase* (ACS) and an inorganic pyrophosphatase (*PPase*) in Methanosaeta species. The results obtained here reveal that these enzymes are the outcome of HGT. The phylogenetic distribution of the ackA/pta system between archaea and bacteria further show that such system is an exclusive bacterial acquisition in Methanosarcina species, whereas the direction of the transfer for the ACS and PPase genes in these methanogens is unclear.

On the one hand, ACS is present in all species of the groups Methanosaeta and Others exhibiting a widespread distribution across methanogens. However, ACS is absent in the species of Methanosarcina analyzed here. It is important to note that the homologous counterparts of ACS were identified by using the reference gene MTHE_1194, which has been reported to perform with high enzymatic activity in Methanosaeta species under methanogenic conditions (Berger et al. 2012). However, the several copies of MTHE_1194 were also identified across Bacteria and Archaea and clustered in diverse protein families, so that the tree topologies of such families show complex scenarios of multiple bacterial gene transfers (see Table 7.4). Interestingly, the gene copies of MTHE_1194 described with poor enzymatic activity in the presence of acetate by Bergen et al. (2012), were classified in other protein families as archaeal specific rather than horizontally transfers. On the other hand, genes of the carbon monoxide dehydrogenase/acetyl-CoA decarbonylase complex (CODH/ACS), which performs the next reaction in this pathway after the acetate activation, were identified as HGT events exported from Archaea to Bacteria. Archaeal homologs of the CODH/ACS in Methanosarcinales families are not exclusive of methanogens (see lower panel the Figure 7.10). Similarly, the subunit mtrH of the methyltetrahydromethanopterin coenzyme M methyltransferase was identified as an export, while the subunit mtrA, defined as archaeal specific, has a paralog classified as HGT.

Figure 7.11 shows the gene distribution of 12 reference genes describing the **methylotrophic methanogenesis** across Methanosarcinales and other methanogens**.** A total of 11 reference genes are involved in HGT events, three of which were classified as imports (*mtmB*, *mttB*, *mttC*), six HGTs were recognized as exports (*mtmC*, *mtbB*, *mtaA, mtaB, mtaC*, and *mtsA*), and the direction of two HGT events were not possible to be defined (*mtbC* and *mtbA*). From the set of genes *mtsA* and *mtsB*, only *mtsB* was defined as archaeal specific. Two major patterns can be observed from the gene distribution and the HGT categories per gene. First, the distribution of methyltransferases of this pathway is highly restricted within Archaea, in Methanosarcinales (7 species) and Methanobacteriales (4 species): *Methanobacterium sp*. (strains SWAN-1 and AL-21),

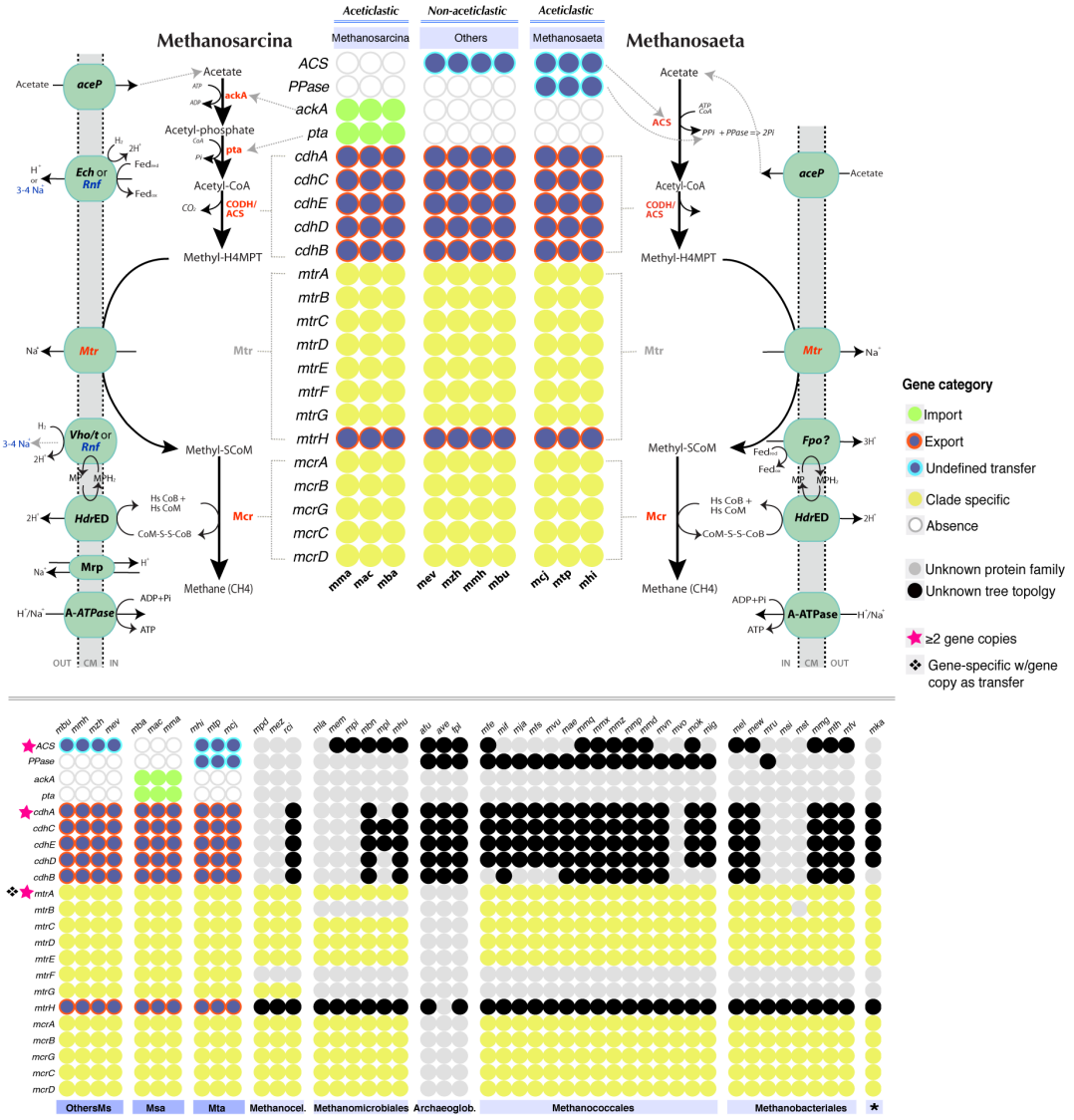


Figure 7.10. Distribution of genes involved in the aceticlastic methanogenesis and their classficiation according to the presence or absence of HGT in the gene family where they belong. The presence (filled circles) or absence (empty circles) of genes of this pathway is shown for the 10 Methanosarcinales genomes (upper figure), as well as their distribution across other methanogenic groups (lower figure). Methanosarcinales species are divided in groups as followed: **Methanosaeta**: *M. harundinacea* (mhi), *M. thermophile* (mtp), *M. concilii* (mcj); **Methanosarcina**: *M. barkeri* (Fusaro) (mba), *M. acetivorans* (mac), *M. mazei* (mma); members of the group **Others**, *M. burtonii* (mbu), *M. mahii* (mmh), *M. zhilinae* (mzh), *M. evestigatum* (mve). Gene categories depict the acquisition from bacteria or import (green), the transfer from archaea to bacteria or export (outlined circle in red), and gene transfers where its direction was undefined (blue). Genes present in the gene family, but the monophyly was not tested (black circle), and genes absent in the family, but probably present in other family with no Methanosarcinales proteins are shown in the lower figure. Finally, genes with no bacterial homologs (archaeal-specific) are shown in yellow circles. **Upper figure:** core genes of both acetate activation, in Methanosarcina (left side) and Methanosaeta (right side), as well as their corresponding schematic reactions of the pathway are shown. **Lower figure:** Gene copies are shown at left side (stars colored in pink), while genes classified as clade specific but with copies classified as HGT events are shown as diamonds. The asterisk (*) represents the archaeon from the Methanopyrales group. Gene distribution of subunits in each membrane model in methanogens is shown in Figure 7.7. The number of homologous proteins out of the Methanosarcinales families (i.e. 'unkown protein family') was estimated only for 48 genes, and is shown in the Supplementary Table ST2. The species names and gene categories description applies also for the Figures 7.11 and 7.12.

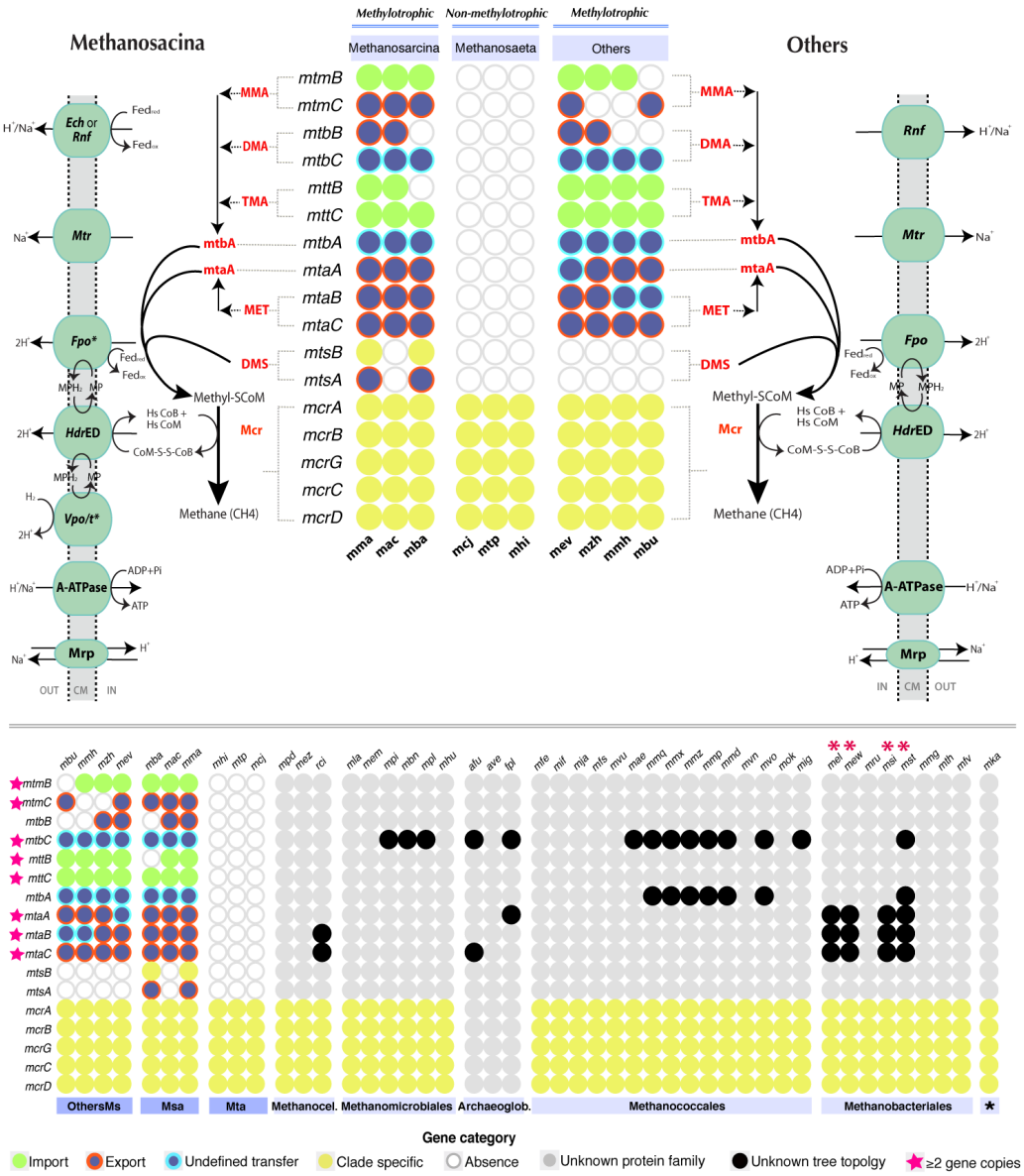


Figure 7.11. Distribution of genes involved in the methylotrophic methanogenesis and their classficiation according to the presence or absence of HGT in the gene family where they belong. Distribution of methyltransferases (MTs) involved in the initial steps of this pathway across Methanosarcinales (upper figure) and other methanogenic archaea (lower figure). **Upper Figure:** The Non-methylotrophic (Methanosaeta), and methylotrophic (Methanosarcina and Others) methanogens are shown at center. Gene categories, as shown in Figure 7.10, depict the presence (filled circle) or absence (empty circle) of a gene, and the colors represent the presence or absence of HGTs (see "Gene category" at the bottom). **Lower Figure:** Gene distribution across other methanogens by their presence in the gene family, but the monophyly was not tested (black circles), and and genes absent in the family, but probably present in other family with no Methanosarcinales proteins (grey circles). Gene copies are shown at left side (stars colored in pink). The asterisk (*) represents the archaeon from the Methanopyrales group. The presence-absence of models have been drawn based on findings of this thesis and supported by other reports: Meuer et al. (2002), Hovey et al. (2005), Li et al. (2006), Smith and Ingram-Smith (2007), Rohlin and Gunsalus (2010), Schlegel et al. (2012a), Schlegel et al. (2012b), Mayer and Müller (2013), Welte and Deppenmeier (2011), Welte and Deppenmeier (2013). The number of homologous proteins out of the Methanosarcinales families (i.e. 'unkown protein family') was estimated only for 48 genes, and is shown in the Supplementary Table ST2. Methylotrophic core genes (*mtmBC, mtbBC, mttBC, mtbA, mtaABC and mtsAB*) and membrane models for energy-conserving electron (left and right sides of the figure) are shown.

Table 7.4. Classification of HGT events as imports (I), exports (E) and undefined (U), and genes with no HGT as specific (S), according to their gene families and methanogenic pathway.

Pathway	*Substrate*	*Gene*	*Total*	*S*	*I*	*E*	*U*	*Gene family*
Aceticlastic	acetate	*ACS*	2	1			1	23374 | 315
		ackA	1		1			17768
		pta	1		1			14733
		cdhA	2			1	1	1187 | 6401
		cdhB	1			1		1416
		cdhC	1			1		1291
		cdhD	1		1			1321
		cdhE	1		1			1296
Methylotrophic	monomethylamine	*mtmB*	2	1	1			14531 | 7632
		mtmC	3	1		2		10275 | 11631 | 23616
		mtbB	1			1		10248
	dimethylamine	*mtbC*	3	1	1		1	12321 | 1444 | 19580
		mttB	3		2		1	13840 | 23064 | 7761
		mttC	2	1	1			17318 | 6586
		mtbA	1				1	4548
	methanol	*mtaB1*	4	2		1	1	10092 | 21653 | 23516 | 5240
		mtaC	3	1		2		12558 | 6268 | 6754
		mtaA	5	3		1	1	10194 | 13606 | 13632 | 14207 | 4984
	dimethylsulfide	*mtsB*	1	1				24166
		mtsA	1			1		24536
hydrogenotrophic	CO2/H2/Formate	*fwdA, fmdA*	2	1	1			3390 | 905
		fwdB ,fmdB	3	1		1	1	1906 | 5200 | 885
		fwdC, fmdC	4	3			1	15636 | 2337 | 3491 | 793
		fwdD, fmdD	3	2		1		1001 | 6342 | 6351
		fwdE, fmdE	2			1	1	1639 | 758
		fwdF, fmdF	2	1		1		6314 | 991
		fwdG	2	1		1		13336 | 4951
		ftr	1		1			1078
		mch	1		1			659
		mtd	1	1				1077
		frhA	1				1	935
		frhB	4	1		1	2	1481 | 3366 | 9331 | 937
		frhD	1				1	5518
		frhG	1			1		996
		mer	1		1			678
core reaction	core enzyme	*mtrA2,mtrA*	3	2		1		1188 | 6248 | 9904
		mtrB2,mtrB	1	1				1385
		mtrC	1	1				1169
		mtrD	1	1				1181
		mtrE	1	1				1186
		mtrF	1	1				6585
		mtrG	1	1				5155
		mtrH	1				1	739
		mcrA	1	1				871
		mcrB	1	1				907
		mcrC	1	1				1191
		mcrD	1	1				850
		mcrG	1	1				870

Gene family category: *S, Specific; I, Import; E, Export; U, Undefined transfer*

Methanobrevibacter smithii ATCC-35061, and *Methanosphaera stadtmanae* DSM_3091). Noteworthy, Methanosarcinales have a major diversity of enzymes to process different methylated substrates when compared to Methanobacteriales, which are only capable to use methanol. Second, there is a patchy distribution of HGTs categories in several of the genes encoding coupled reactions, such as that observed in methanol (*mtaA*, *mtaB*, *mtaC*), monomethylamine (*mtmB*, *mtmC*) and dimethylamine (*mtbA*, *mtbB*, *mtbC*). Hence, a specific pattern of bacterial acquisitions for a particular pathway cannot be observed. Within monomethylamines, for instance, one gene was apparently acquired from bacteria (*mtmB*), while the other (*mtmC*) appears to have been exported to bacteria. This heterogeneous pattern is recurrent in genes associated to substrates of

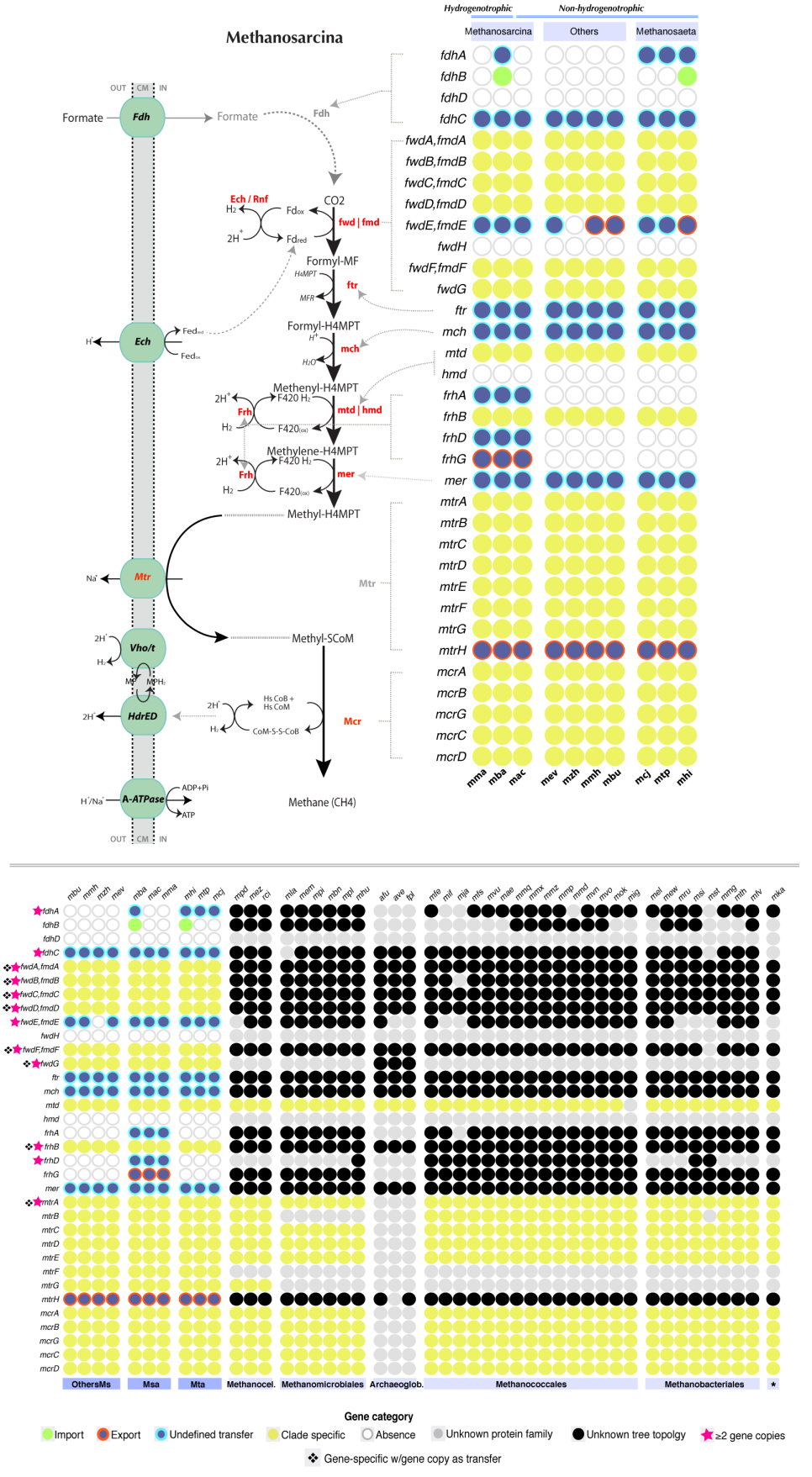


Figure 7.12. Distribution of genes involved in the hydrogenotrophic methanogenesis and their classification according to the presence or absence of HGT in the gene family where they belong (Description continues next page).

Figure 7.12. (Continue from the previous page). Distribution of enzyme involved in the hydrogenotrophic pathway across Methanosarcinales (upper figure) and other methanogenic archaea (lower figure). **Upper Figure:** Hydrogenotrophic methanogens are shown at left side (*M. mazei,* mma; and *M. barkeri,* mba), and the others species represent non-hydrogenotrophic methanogens (*M. acetivorans* (mac), members of Methanosaeta and Others). Hydrogenotrophic core genes and protein complexes collapsed names from top-to-bottom: *fwd/fmd, ftr, mch, mtd/hmd, frh, mer, mtr,* and *mcr*) and membrane models for energy-conserving electron (*fdh*, *ech*, *vho/t*, *hdrDE*) are shown at the left side. Gene categories, as shown in Figure 7.10, depict the presence (filled circle) or absence (empty circle) of a gene, and the colors represent the presence or absence of HGTs (see "Gene category" at the bottom). **Lower Figure:** Gene distribution across other methanogens by their presence in the gene family, but the monophyly was not tested (black circles), and and genes absent in the family, but probably present in other family with no Methanosarcinales proteins (grey circles). Gene copies are shown at left side (stars colored in pink). The asterisk at the bottom of archaeal clades (*) represents the archaeon *Methanopyrus kandleri*. The presence-absence of models have been drawn based on findings of this thesis and supported by other reports: (Deppenmeier 2002), (Welander and Metcalf 2005), (Thauer et al. 2008), (Kulkarni et al. 2009). The number of homologous proteins of 'unkown protein family' category was estimated only for 48 genes, and is shown in the Supplementary Table ST2.

this pathway, except those related to dimethylsulfide (*mtsA*, *mtsB*). A manual inspection of the corresponding gene trees, followed by an extensive searching of these proteins across all prokaryotic genomes, confirmed their limited presence in archaea and bacteria, but a considerable number of gene copies were found in most of them (see starts colored pink at the lower panel in Figure 7.11 and Table 7.4).

Finally, Figure 7.12 shows the distribution analysis performed on the reference genes of the **hydrogenotrophic pathway**, which reveals three major results: 1) its gene repertory has a high number of paralogs; 2) some of its genes are associated to HGTs; and 3) the gene complex F_{420}-reducing hydrogenase (*frhA, frhD, frhG, frhB*), key for the oxidation/reduction of the F_{420} coenzyme of this pathway, is absent (probably loss) in some Methanosarcinales (see upper panel at the Figure 7.12). By following the direction of the chemical reactions at upper panel in Figure 7.12, it is possible to observe that five (out of the seven) of the genes playing a crucial role in the core reactions of this pathway are in a single version (*ftr*, *mch*, *mtd* and *mer*), while most subunits of the protein complexes have ≥2 copies, such as *fwdA, fwdB, fwdC, fwdD, fwdF, fwdG*, *frhB*, *mtrH* and *mtrA* (see diamonds next to the gene names at the lower panel in Figure 7.12 and Table 7.4). From those subunits with copies, one version has no bacterial homologs (i.e. this is archaeal specific), whereas the other copies are associated to HGT events, mainly as exports (after a manual inspection). In addition, 8 enzymes were classified as undefined transfers, including *fwdE*, *ftr*, *mch*, *frhA, frhD, mer, mtrA* and *mtrH* (the latest two proteins were described in the previous aceticlastic pathway section). The archaeal homologs of this set of enzymes are widely spread in almost all methanogens; however, some subunits, such as *fwdG*, *fwdF*, *frhA, frhD, frhG*, and *mtrB, mtrF, mtrG,* were not identified in the gene families of Methanosarcinales (see lower panel in Figure 7.12). Similarly, only Methanosarcina species and other methanogens with cytochromes contain the cytoplasmic F_{420}-reducing hydrogenase (*frhA, frhD, frhG, frhB*), while is absent in the other Methanosarcinales species (see upper panel in Figure 7.12).

Based on the consensus of 10 Methanosarcinales genomes, multiple HGT events were detected and related to key steps in methanogenesis and energy production. The results described previously suggest that key HGT events were involved in an expansion of the aceticlastic pathway and probably also in the origin of the methylotrophic pathway. Furthermore, the absence of the Frh hydrogenase

identified only in *substrate-specific* methanogens might be related to the incapability of these organisms to grow in hydrogenotrophic conditions. These evolutionary inferences might be, however, strongly biased due the limited number of genomes analyzed in this study (NCBI database version 2012), despite the fact that the 10 analyzed genomes still represent a good diversity of the increased number of sequenced genomes from the Methanosarcinales clade.

7.8. Experimental reports and an updated analysis of Methanosarcinales genomes suggest the genomic expansion and loss events in methanogenesis.

To test the robustness of the previous inferences in Methanosarcinales, a further *comparative genomics* analysis and a *survey in the literature* of physiological features of these species were performed for 24 new complete genomes (version March 2017). Panel C in the Figure 7.13. As described above, major gene changes (i.e. gains and losses genes) occurred at the membrane level, while keys bacterial acquisitions via HGT are located at the first step of the aceticlastic and methylotrophic pathways (see red asterisk and blue crosses symbols). On panel A in Figure 7.13, the comparative genomics analysis over 34 Methanosarcinales genomes display the identification of orthologous genes related to methane and energy metabolism. The detection of orthologous genes was also supported by the identification of the same functional PFAM domains detected in the reference proteins (see Pfam domains designations in Figure 7.8 and pairwise comparison of global identity percentages Figure 7.14). The results exhibit a large number of gene gains and losses (see cross symbols colored in blue on Figure 7.13) in the electron transport chain and also in the aceticlastic and methylotrophic pathways. Examples for the former include protein complexes such as Rnf, Ech, Mpr, F-ATPase, Vho/Vht, and Frh, and genes for the methanogenic pathways include genes such as latter such as the ackA/pta system, ACS, membrane permeases, and some methyltransferases for dimethylsulfide (*mtsA*, *mtsB*, *mtsC*). Some of these HGT events were clearly identified as bacterial acquisitions (red colored asterisk symbols), such as Rnf, Mpr, F-ATPase, ackA/pta system, ACS, biosynthesis of the cofactor F_{420} (*cofC*, *cofD*, *cofE*), pyrrolysine (Pyl operon), and all the permeases related to the uptake of methanogenic substrates. Other HGT events (those marked with asterisk symbols colored in cyan on Figure 7.13) were classified as ambiguous, since it is not possible to identify them as bacterial acquisitions or archaeal exports.

Panel B in Figure 7.13 displays the wide range of physiological conditions on which 48 Methanosarcinales (species and strains) can grow and perform methanogenesis. The analysis of experimental reports shows that most Methanosarcinales species have preference for one or a few substrates, but they are not capable to perform all methanogenic pathways. This trend is clear in the aceticlastic Methanosaeta species and in multiple methylotrophic methanogens, especially in the "obligated" methylotrophic Methanosarcinales. The experimental reports compile analysis of 16 out of 25 species from the genus Methanosarcina, which show the flowing trends: 1) only 6 species with a freshwater lifestyle use more than one type of substrate and methanogenic pathway (see red squares on Figure 7.13), and 2) the remaining species of this genus can only perform either the aceticlastic or the methylotrophic methanogenesis. In addition, from the 16 species with experimental reports, 12 strains were tested to growth under hydrogenotrophic conditions, where 5 grow and 7 do not grow. Only 2 species have difficulties of growing under hydrogenotrophic or aceticlastic conditions, such as *M. mazei* S6, *M. thermophila* TM1, *M. baltica* and *M. siciliae* T4M (see circles in cyan color on Figure 7.13). And yet, 16 out of 25 exanimated Methanosarcina species in the literature have not been tested for all substrates, which leaves an open question about their capability to grow on diverse conditions.

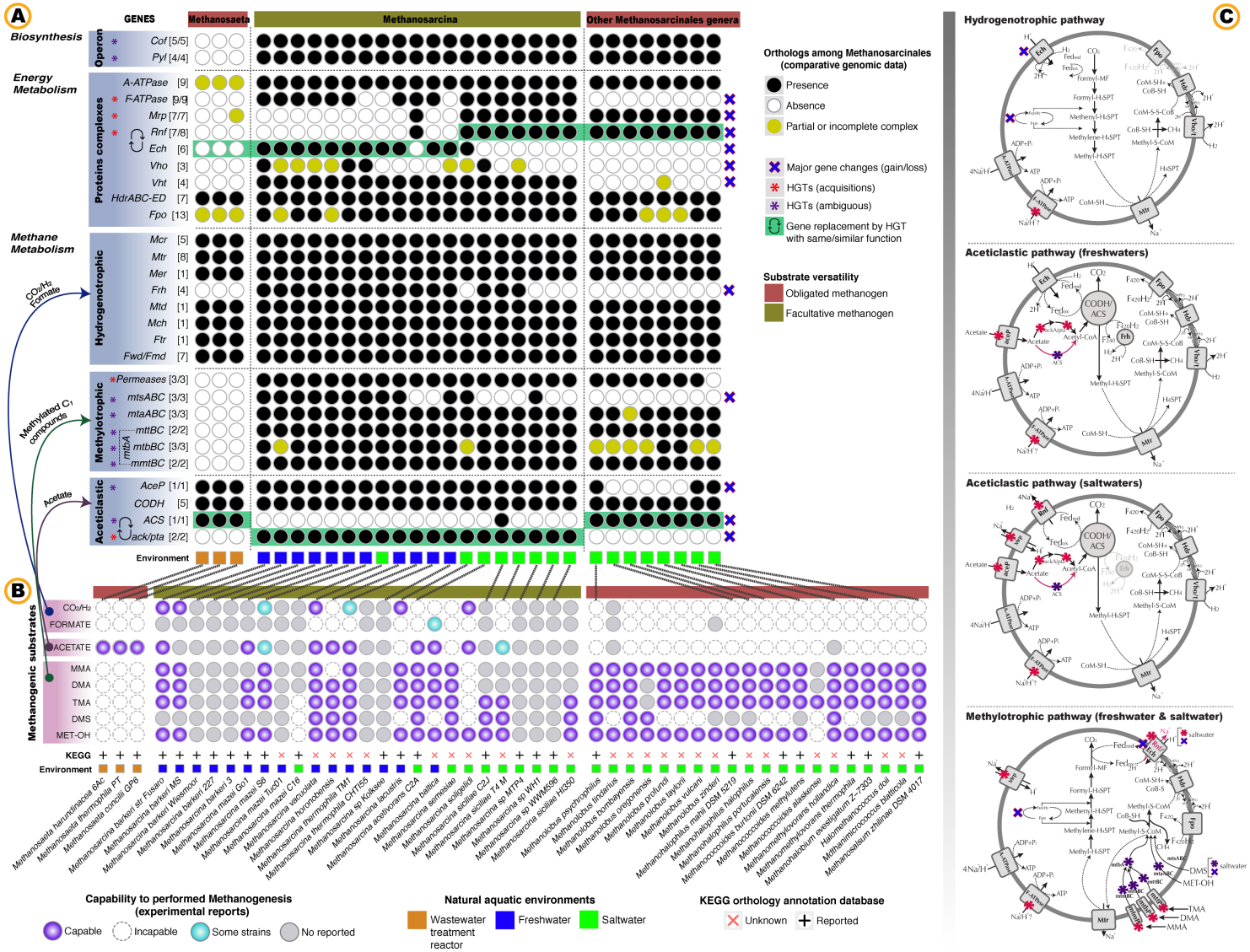


Figure 7.13. Capability to perform methanogenesis: experimental and genomic data in Methanosarcinales through different pathways (Continues next page).

Figure 7.13. The capability of Methanosarcinales to perform methanogenesis with different substrates (Continue from last page). The graphical summary (panel C) of the genomic findings in this study (panel A), and examined experimental data (panel B) is shown from the analysis of 48 Methanosarcinales species and strains. Methanosarcinales species were classified by major taxa and substrate dependency as: Methanosaeta (obligated aceticlastic), Methanosarcina (substrate facultative) and other Methanosarcinales genera (obligated methylotrophic). **Panel A.** The presence (circles in black) or absence (empty circles) of orthologs across 34 complete genomes is shown from 84 functional reference genes (see Methods) divided in three categories: biosynthesis (biosynthesis of cofactor420 and pyrrolysine), energy metabolism (several membrane protein complexes), and methane metabolism (methanogenic pathways: hydrogenotrophic, methylotrophic and aceticlastic). Abreaviated names (genes and complexes) are shown close to box brackets ([ ]), it enclose their total number of genes (number at left side), and the total likely bacterial acquisitions (number at right side). Circles in yellow depict incomplete protein complexes or set of methyltransferases belonging to a specific function o pathway, respectively. HGT events were estimated in 10 species (asterisks), and depict bacterial acquisitions (red color) or ambiguous (blue color) depending on the direction of the transfer. Major gene changes (gain/loss events) are highlighted with a cross colored in blue (right side). Functional replacement of one protein/complex by another is shown with recycling arrows (added green color behind circles). Individual gene scores for global identity and coverage percentage are shown in the Figure 7.14, and Supplementary Figure SF7, respectively. **Panel B.** The growth capability (circle in purple) or incapability (empty circle) of 48 Methanosarcinales species and strains in specific substrates, as compiled from the literature. Circles in blue color depict the restriction of some strains to grow in the corresponding substrate. Circles in grey color represent species with no experimental report yet. Symbology of substrate names by methanogenic pathway, hydrogenotrophic: CO_2/H_2 and/or formate; aceticlastic: acetate; and methylotrophic: mono-,di-,trimethylamine (MMA, DMA, TMA, respectively), dimethylsulfide (DMS), and methanol (MET-OH). Gene functions were obtained from the KEGG, when available (+) and no available (x). The natural habitat of the species analyzed, as reported from the experiments or from the BacDive database (Söhngen et al., 2016), includes saltwater (squares in green color), freshwater (squares in red), and wastewater treatment (squares in yellow). Complete forty-eight to forty-eight Methanosarcinales species comparison of genomic and experimental findings is shown is Supplementary Figure SF8. **Panel C.** Location of the HGTs in the methanogenic pathways and in the electron chain transfer in Methanosarcinales. At the membrane level, metabolic changes involve genes in the uptake of methanogenic substrates (AceP and other permeases), as well as in electron transfer chain (Ech and Rnf) and ion H+/Na+ pumping (Mrp, F-ATPase).

7.9. Osmoregulation of halotolerance genes in Methanosarcinales is related to HGT events with bacteria.

Methanosarcinales species are considered as either facultative or as extreme-moderate halophiles. Some of these species use the *salt-out* mechanism to live in this environment, which implies the accumulation of compatible solutes by uptake or *de novo* synthesis using specific enzymes. However, the wide presence of this mechanism in Methanosarcinales and its putative origin has not been investigated yet, with the exception of the nonhalophilic *M. mazei* (Pflüger et al. 2007). Thus, 94 reference genes involved in osmotic pressure and salt adaptation from M. *mazei* were screened across all Methanosarcinales genomes to find orthologous proteins and HGT events. The results from this analysis are observed on Figure 7.15 and supplementary Figure SF9.

From the 94 reference genes, only the homologous counterparts of 85 genes were detected, and from these, 57 protein genes were identified as HGT events. The remaining 28 protein genes were classified as archaeal specific, 18 genes of which occur exclusively in Methanosarcinales, while the remaining 10 genes are shared between Methanosarcinales and other archaea. From the total 57 cases of HGT events, 4 were classified as exports, 14 as imports, and 39 as undefined transfers (see Figure 7.15). The functions of the identified transfers include genes such as transport and binding sites (ABC transporter, permeases, cobalt and phosphate transporters (*pstB*, *pstC*, *pstS*)), stress and regulation (conserved and hypothetical proteins), and other enzyme with unknown functions. On

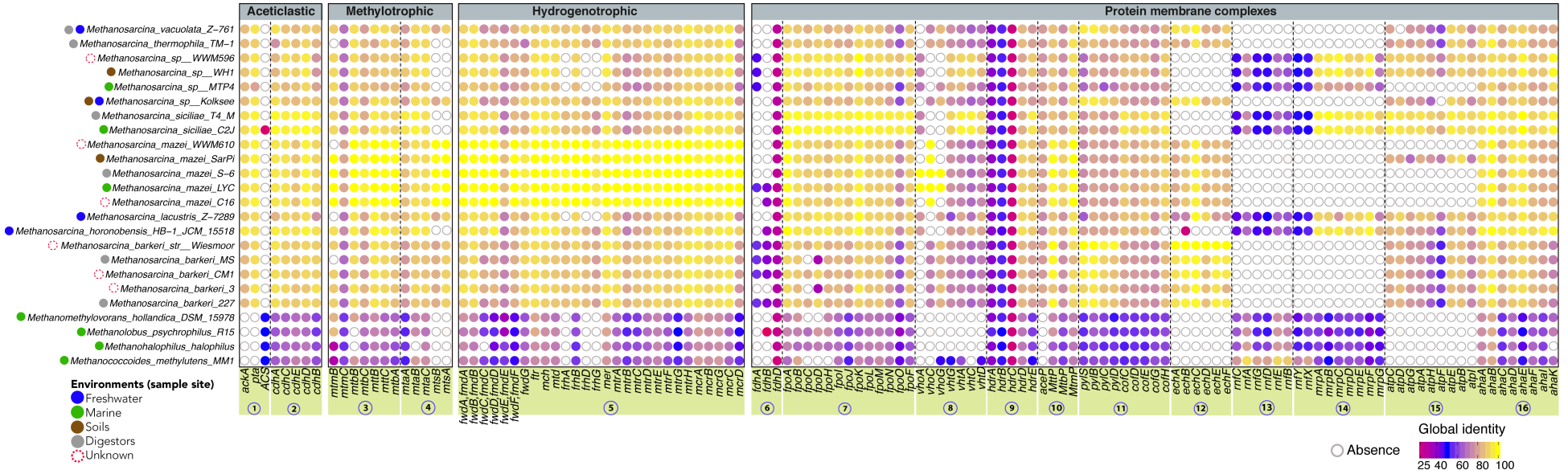


Figure 7.14. Orthologous genes between genomes of Methanosarcinales involved in methane and energy metabolism with their corresponding pairwise score of global identity. Presence (filled circles) and absence (empty circles) of orthologous genes between members of the family Methanosarcinaceae. Parwise score of global identify in percentage (≥25%) is depicted by color intensity (see scaled bar at rigth). Four functional categories are shown (at top): methanogenesis (aceticlastic,1-2; methylotrophic, 3-4; hydrogenotrophic, 5), and multiple proteins complexes (associated to the membrane, 6, 7, 8, 9,10, 12-16; biosynthesis of the cofactor$_{430}$ and pyrrolysine, 11). Species of the Methanosarcinaceae are shown at left side. Environmental lifestyle was obtained from the literature or via the BacDive database (Söhngen et al. 2016). Proteins and protein membranes description is as follow (from left to right): 1, genes in aceticlastic methanogenesis exclusively present in Methanosarcina species; 2, genes in aceticlastic methanogenesis exclusively present Methanosaeta species; 3 genes involved with mono-,di-,tri-methylamines as substrates; 4, genes involved with methanol (mta) and dimethylsulfide (mts); 5, genes involved in the hydrogenotrophic methanogenesis; 6, formate incorporation; 7, the $F_{420}H_2$ dehydrogenase complex (fpo); 8, methanophenazine-reducing hydrogenase (vht/vho); 9, heterodisulfide reductase (hdr); 10, membrane transporters for acetate (aceP) and methylated compounds (mttP, mtbP, mmP); 11, genes part of the pyrrolysine operon (pyl); 12, Ech hydrogenase (ech); 13; Rnf complex (rnf); 14, Na^+/H^+ antiporter (Mrp); 15, bacterial F-type ATPase (atp); 16, archaeal A-type ATPase (aha). Global pairwise alignments were carried out for 24 Methanosarcinales genomes comparing the reference genes (see Tables 5.2, 5.3, 5.4, and 5.5 in Methods) against all proteins using the Needleman-Wunsch algorithm. The presence (filled circles) or absence (empty circles) of a gene is shown. The percentage of sequence coverage comparing each sequence query against each target sequence is shown in the Supplementary Figure SF7.

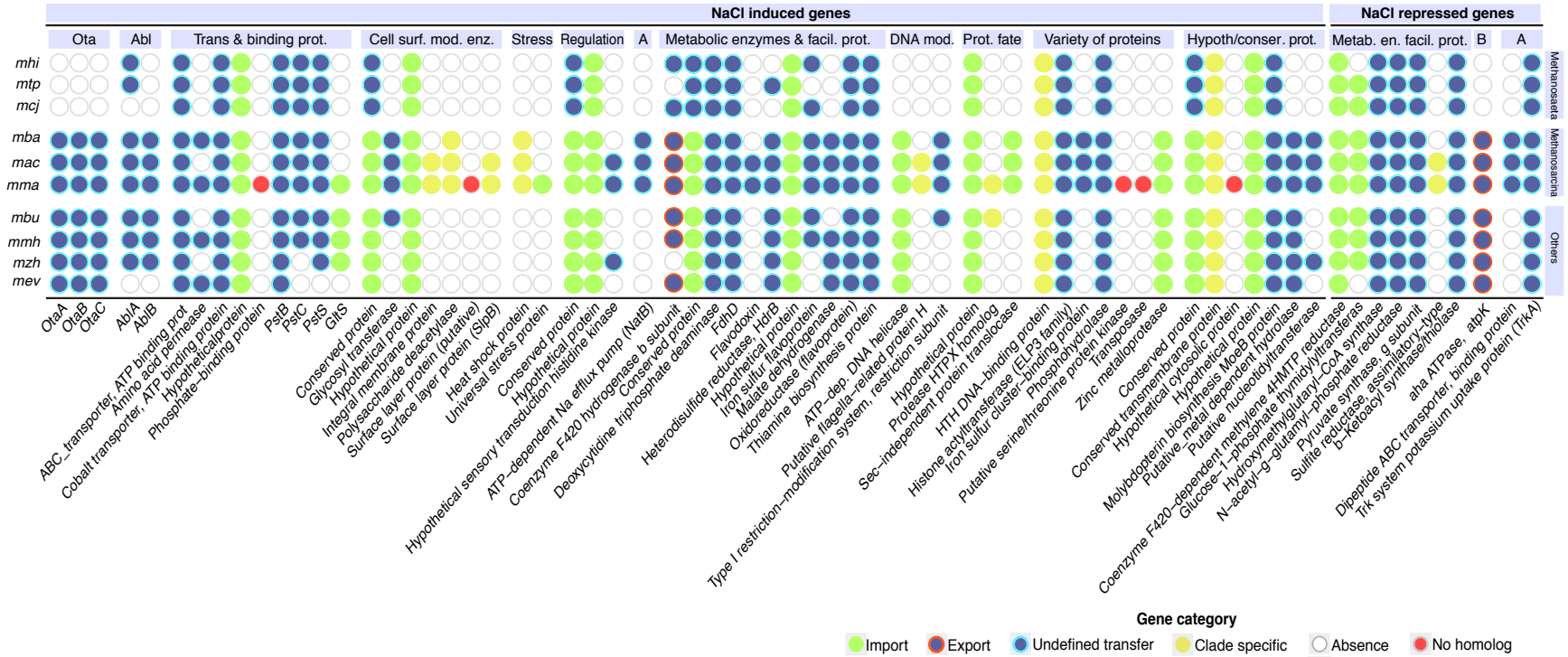


Figure 7.13. Orthologus genes between Methanosarcinales species related to the salt adaptation and their classification for HGT. Distribution of orthologous genes between *M. mazei* and other Methanosarcinales distributed in 85 gene families. Genes were expressed (NaCl induced genes) or repressed (NaCl repressed genes) during growth in saline conditions, from moderate to high conditions, as was reported by Pflüger et al. (2007). Gene categories depict the presence (filled circle)—as import (green), export (red), and undefined transfer (blue)—or abence as archaeal-specific origin (yellow) of the HGT. Emtpy circles represent absence of the gene. Major functional categories are described as followed (from left-to-right): *NaCl induced genes;* glicyne-betaine transport, biosynthesis of Ne-acetyl-lysine, transprot and binding proteins, cell surface modifying enzymes, stress reponse, regulatory functions, Na exporter, Metabolic enzymes, DNA modifying proteins, protein degradation and secretion, putative or hypothetical proteins; *NaCl repressed genes*, metabolic enzymes, transport and binding proteins. Corresponding gene names are shown at the bottom. Methanosarcinales group: *Methanosaeta* (mhi, mtp, mcj), *Methanosarcina* (mba, mac, mma) and *Others* (mbu, mmh, mzh, mve). Methanosarcina species code names: **Methanosaeta (Mta)**, mhi, *M. harundinacea*; mtp, *M. thermophile*; mcj, *M. concilii*; mba; **Methanosarcina (Msa)**, *M. barkeri* (*Fusaro*); mac, *M. acetivorans*; mma, *M. mazei*; **Others (Othrs)**, mbu, *M. burtonii*; mmh, *M. mahii*; mzh, M. zhilinae; mve, *M. evestigatum*. A complete distribution of the homologs of this set of halophilic adaptation system among all species in archaea is shown in the Supplementary Figure SF9.

the other hand genes with no bacterial homologs are related to cell surface and several conserved proteins with hypothetical functions, such as glycosyl transferase, surface layer proteins, transmembrane proteins or DNA binding sites. Interestingly, there are reference genes, which appear to be exclusive of *M. mazei*, with no archaeal or bacterial homologs detected (circles in red color on Figure 7.15). These genes were characterized as polysaccharide deacetylase, serine/treonine kinase, transposase, and as cytosolic conserved proteins. After a manual inspection of the 39 undefined transfers, only 7 genes remains with an unclear origin, while the tree topology and density of species suggest that the 32 of these genes were very likely acquired from bacteria.

7.10. Discussion of the chapter: HGT and its biological impact in Methanosarcinales.

The present study has characterized the presence of horizontal gene transfer events across all sequenced members of Methanosarcinales, which supports the view that bacterial genomes are extremely dynamic in acquiring novel capabilities from foreign genes. The findings of this thesis also suggest a differential gene loss and secondary acquisitions that might have occurred after the divergence of the last common ancestor of Methanosarcinales. This is in agreement with our initial report (Nelson-Sathi et al. 2015), as described in Chapter 6 and 7, indicating that some bacterial gene transfers are specifically involved in the origin of the Methanosarcinales clade (imports-MS), while another large fraction of bacterial acquisitions (imports-Archaea) were possibly acquired by multiple and independent HGT events. The present work also assessed the presence of the HGT in genes related to the osmoregulation of halotolerance, and in specific pathways of the methane and energy metabolisms of Methanosarcinales. The findings presented here show that HGT events occur in multiple genes of these biological processes, suggesting that this evolutionary mechanism have also promoted (to some extent) the metabolic versatility in Methanosarcinales. Throughout the present section, I will discuss the main findings of this thesis, and contrasting them with evidence from literature, to argue about the impact that HGT have had on the ecological and metabolic evolution of Methanosarcinales.

Several reports have previously analyzed the presence of HGTs in some Methanosarcinales genomes by using similar phylogenetic approaches (Deppenmeier et al. 2002; Garushyants et al. 2015; Nelson-Sathi et al. 2012; Youngblut et al. 2015). Most of these studies have focused only, however, on Methanosarcina species due to a limited number of available genomes at that time and their metabolic versatility to adapt to new environments. In agreement with these reports, the findings of this work show that a high number of genes are involved in HGT events, representing between the 9% and 14% of the Methanosarcinales genomes, and a considerable number of them were acquired from bacteria. The study of (Garushyants et al. 2015) found a minor number of bacterial acquisitions in Methanosarcina genomes. The difference could be largely explained by the number of species used in both studies, which is threefold higher in the present work (by including also other archaea) in comparison of the members of species used in the study of Garushyants et al. (2015). Yet, both estimations are very similar if the analysis takes into account species only from Methanosarcina species. For instance, the results found here predicted 246, 220, and 179 imports for *M. acetivorans, M. barkeri*, and *M. mazei*, respectively, while (Garushyants et al. 2015) estimated 221, 214, and 151 imports for the same species.

Moreover, the imports acquired by Methanosarcinales exclusively are found to have bacterial donors mainly from clades belonging to Proteobacteria and Firmicutes. Also, most of the acquisitions identified are associated to metabolic functions rather than to informational roles, which is in

agreement with previous studies (Allen et al. 2009; Deppenmeier et al. 2002; Garushyants et al. 2015). Noteworthy, ribosomal proteins, RNA transport and biosynthesis of tRNAs are also among the HGTs involved in informational functions, which further challenge the hypothesis that conserved and crucial informational genes are unlikely to be laterally transferred (Jain et al. 1999). From these genes, HGT ribosomal proteins might have the highest impact on phylogenetic reconstruction, since these are crucial tools to make biological or ecological inferences among multiple species (reference phylogenetic tree). Therefore, it is becoming important to perform additional research to confirm and categorize the presence of HGT in universal gene markers, which have not only been reported in Bacteria (Brochier et al. 2000; Garcia-Vallvé et al. 2002; Makarova et al. 2001), but that habev been also suggested to be the outcome of ancient transfers (Yutin et al. 2012).

7.10.1. Halophilic and non-halophilic Methanosarcinales share bacterial-like genes for adaptation to saline environments.

Several studies have suggested that most halophilic archaea use the *salt-out* system (which accumulate similar compatible solutes after its biosynthesis) to cope with external salinity (Chaban et al. 2005; Spanheimer and Müller 2008). The exceptions of this trend are members of haloarchaea (Oren 2002) and species from the recently reported group named 'Methanonatronarchaeia' (Sorokin et al. 2017), which use the *salt-in* system promoting intracellular accumulation of organic 'compatible' solutes. Yet, it remains largely unclear the origin and evolution of such systems in methanogens and, especially, in halophilic methanogens such as Methanosarcinales. In the present work, thus, 84 reference genes from *Methanosarcina mazei* involved in coping with high salinity in freshwater enviroments (Pflüger et al. 2007) were used to identify homologous counterparts within Methanosarcinales and other methanogens, such as *Methanocelalles, Methanomicrobiales,* and *Methanococcales*.

The findings obtained here show that several homologous proteins of these reference genes are found in almost all members of the family *Methanosarcinaceae* (covering both freshwater and marine species). However, only a few homologs were detected in the members of *Methanosaetaceae*. I also found that most of the homologous genes identified are involved in HGT events, and quite probably were also acquired from bacteria through multiple transfers. Nevertheless, further and extensive analyses including more halophilic methanogens are necessary to be conclusive about the phylogenetic identity of the donors, as well as regarding the ancestry of these HGT events in Methanosarcinales. It is important to note that the bacterial acquisitions found in members of *Methanosarcinaceae* involve three gene sets related to salt adaptation: *i*) biosynthesis and transport of compatible solute (salt-out system), *ii*) regulatory proteins, and *iii*) a diverse set of other metabolic proteins such as iron sulfur, malate dehydrogenase, biosynthesis of thyamine, deoxicytine triphosphate deaminase, heterodisulfide reductase (hdrE), formate dehydrogenase (fdhD) and others. The presence of genes from the salt-out system, such as Abl and Ota, in other methanogens supports the common use of similar solutes to cope high environmental salinities, as suggested in previous studies (Enpadinhas and da Costa 2008; Lai and Gunsalus 1992; Lai et al. 2000; Lai et al. 1991; Roeβler and Müller 2001; Sowers and Gunsalus 1995).

7.10.2. A set of bacterial genes involved in energy metabolism and acquired by HGT led to an adaptation of methanogens to saltwater environments.

It has been largely acknowledged that members of the family *Methanosarcinaceae* can be "divided" in methanogens that inhabit either freshwater environments and use H_2 or saltwater environments and not use H_2 as an intermediate molecule in methanogenic pathways (Allen et al. 2009; Ferry and Lessner 2008; Guss et al. 2009; Kulkarni et al. 2009; Spring et al. 2010; Wang et al. 2014; Welte and Deppenmeier 2013). To accomplish such environmental adjustments, a varied set of membrane-bound protein complexes in Methanosarcinales are involved in the chemiosmotic gradient that can be utilized to form energy (ATP) by an ATP synthase complex. However, no all methanogens use the same system of membrane-bound protein complexes for this process. In agreement with this, my results show the presence of four conserved membrane-bound protein complexes in all Methanosarcinales species, whereas other protein complexes are exclusively present either in freshwater or in saltwater species. For example, the four membrane complexes conserved in all Methanosarcinales are the two versions of heterodisulfide reductase (HdrED and HdrABC), the $F_{420}H_2$ dehydrogenase (Fpo), and the A-type ATPase synthase (Aha). On the one hand, the freshwater species only harbor the two versions of the methanophenazine-reducing hydrogenase (Vho and Vht) and the energy conserving hydrogenase (Ech), whereas the saltwater species harbor the the F-type ATPase synthase (Atp), the Na^+/H^+ antiporter (Mrp), and the Na^+ transporting NADH oxidoreductase Rhodobacter nitrogen fixation (Rnf). Noteworthy, the homolooug counterparts of these complexes in Methanosarcinales (with ≥ 50% of identity) also have all functional Pfam protein domains conserved, which further support their functional role in the electron transfer chain.

It is interesting to note the identification of specific HGTs event that divide freshwater and saltwater species within Methanosarcinales. First, it is observed the replacement of the "H_2-evolving Ech hydrogenase" complex by the bacterial ferredoxin oxidation function of the Rnf complex (Ferry and Lessner 2008). This event appears to be followed by secondary acquisitions of Mrp and F-ATPase, the former being an antiporter that optimizes the transmembrane proton gradient to maintain an ideal ATPase (*aha*) function in salty environments (Jasso-Chávez et al. 2013; Jasso-Chávez et al. 2016; Pflüger et al. 2007; Schlegel et al. 2012a; Spring et al. 2010). Remarkably, the findings of the present thesis suggest that the presence of the Rnf, Mrp, and F-ATPase complexes is restricted to Methanosarcinales, particularly in the *Methanosarcinaceae* family, and further support the idea that these genes have been key factors reponsable for the freshwater-saltwater transition in these methanogens (Li et al. 2006; Wang et al. 2011; Welte and Deppenmeier 2013). For instance, the presence of both complexes Ech and Rnf in one Methanosarcina species, *M. lacustris*, suggests a transition between both membrane complexes and salt-fresh waters. On the other hand, the role of the F-ATPase in the adaptation to saline-freshwater environments is unclear. The presence of the F-ATPase has been proposed as *non essential* for the cellular growth on this methanogens, so that, it has became a 'dead-ended' HGT event (Rohlin and Gunsalus 2010). However, there are other studies suggesting its role, as pumping Na^+, in marine environments (Dibrova et al. 2010), as well as a minor unknown role has been suggested in the aceticlastic methanogenesis in some Methanosarcina species (Rohlin and Gunsalus 2010).

Several reports have suggested, but not fully demonstrated, that these protein complexes come from bacteria via HGT events. Two main strategies have been previously used to this end: 1) the identification of close homologs of the Rnf complex via BLAST analysis in two Methanosarcinales species such as in *M. Mazei* (Deppenmeier et al. 2002; Youngblut et al. 2015); or 2) by finding functional and structural similarities between archaeal and bacterial protein complexes (Spanheimer

and Müller 2008; Welte and Deppenmeier 2013). The phylogenetic findings of the present study robustly demonstrate that these membrane-bound protein complexes have been indeed acquired from Bacteria via HGT at the ancestor of the *Methanosarcinaceae* family. Furthemore, the presence of Rnf and Mrp in all *Methanosarcinaceae* members further contradicts the findings of Li et al. (2006), who suggested that Rnf was acquired exclusively at the ancestor of the Methanosarcina species.

Here I propose that most of these HGT events have likely occurred previous to the Methanosarcina ancestor, being probably acquired at the origin of the Methanosarcinales clade. This proposal is supported by four additional observations. First, our previous study has shown that Rnf genes were acquired at the origin of the Methanosarcinales clade, and that they are considered as key component of the metabolic innovation of this group (Nelson-Sathi et al., 2015). Second, it has been previously suggested the fundamental role that the Mrp complex has in saline environments for saltwater species such as Methanosarcina (Jasso-Chávez et al. 2013; Jasso-Chávez et al. 2016). By contrast, the homologous counterparts of Mrp in other hyperthermophilic archaea have been reported to be involved in hydrogen and sulfur metabolisms (Ito et al. 2017; Schut et al. 2012). Third, the origin of the Methanosarcina group has been estimated to be close to the end-Permian extinction, placing this group as the most recent within Methanosarcinales (Rothman et al. 2014), while one of their close relatives might represent the oldest members, such as *Methanosaeta* (supplementary figure SF10). Fourth, the gene distribution of the Rnf and Mrp complexes occurs in species living in saline environments and in one Methanosaeta species (which holds an incomplete version of Mrp). These observations suggest that these complexes, including the bacterial ATPase, were very likely acquired at the origin of the Methanosarcinales clade followed by differential gene loss. Nevertheless, further comparative analyses between marine and freshwater living species are needed to support the origin of these membrane-bound complexes in Methanosarcinales and in other methanogens, as well as to explain the presence of some of these complexes in non-methanogenic groups, such as the uncultivated methane-oxidizing archaea group (ANME) (McGlynn 2017; Wang et al. 2014).

Additional studies are necessary to understand the nature of the HGT events associated to other subunits of the complexes Hdr, Ech, Fpo, A-ATPase, Vho and Vht, which are defined as either imports or undefined transfers. Although a manual inspection of these non-monophyletic gene trees and their species density suggest that some genes were very likely exported to bacteria from archaea (e.g. A-ATPase), the fact is that others genes and protein complexes seem to be a bacterial acquisition, such as Fpo, Ech or Vho/t (except the subunits associated to cytochrome) (see Supplementary Figure SF18 in Apendix II). Within these potentially imported complexes, my findings show that only six specific subunits do not have bacterial homologs or even archaeal homologs, suggesting either an archaeal or unique Methanosarcinales origin. Intriguingly, these cases correspond to genes encoding multiheme *c*-type cytochrome (*hdrE*, *vhoC*, *vhtC*, *rnfY*), and proteins of the Fpo complex that interact with methanogenic substrates reduced coenzyme F_{420} (*fpoF*) and methanophenzine (and *fpoO*), whose origin have been described in other reports (Deppenmeier et al. 2002; Thauer et al. 2008; Welte and Deppenmeier 2013).

7.10.3. Expansions in methanogenesis are influenced by HGT events: on the bacterial origin of the aceticlastic and methylotrophic pathways

Aceticlastic methanogenesis. The phylogenetic evidence shown in this study confirms the presence of the HGT in enzymes involved in the initial steps of this pathway, that is, the transport of acetate into the cell through the membrane, the activation of acetate, and the cleavage of acetyl-CoA. First,

the enzymatic systems that activate acetate, as the one exclusively used by Methanosarcina species (the ackA/pta system) and the other used by Methanosaeta species (ACS and PPase), show a different evolutionary history via HGT. For example, the ackA/pta system is a clear case of a single and exclusive bacterial acquisition by Methanosarcina species, as previously reported (Fournier and Gogarten 2008). On the other hand, the "ACS and PPase" in the Methanosaeta species as well as the permease aceP —the unique enzyme transporting the acetate into the cell by all aceticlastic methanogens (Welte et al. 2014)—, were very likely acquired from bacteria in independent HGT events, followed by divergence and frequent loss events across archaea. This is supported by the presence of methanogens from other clades in the trees analyzed. Also, the diversity of ACS among different methanogenic archaea has been suggested via divergence by gene amplification (i.e., an increase in the number of copies), followed by the selection of particular sequence variations for a higher specificity acetate among other archaea (Smith and Ingram-Smith 2007). The divergence of ACS has been analyzed at the genomic level in *Methanosaeta thermophila* and other archaea, whereas the activity of PPase has been examined in Methanosaeta species only (Berger et al. 2012; Smith and Ingram-Smith 2007).

It is interesting to note that the two different enzymatic systems in charge of activating acetate in the two closely related genera of Methanosarcinales, Methanosaeta (ACS and PPase) and Methanosarcina (the ackA/pta system), were identified as HGT events. Previously, Fournier and Gogarten (2008) identified the acquisition of the ackA/pta system from bacteria and suggested, also in agreement with Rothman et al., (2014), the possible existence of a methanogenic aceticlastic metabolism previous to the origin of the ackA/pta system in Methanosarcina. However, evidence to support this scenario has not been provided so far. The findings in the present work not only suggest that both enzymatic systems were very likely acquired independently from bacteria, but they also support the idea that the ackA/pta system was most likely acquired at the ancestor of Methanosarcina, replacing thus the ACS and PPase system in this group. This hypothesis is supported by the following observations.

First, the genes encoding the first steps of the aceticlastic pathway (i.e. aceP, ACS, and PPase) are distributed in almost all aceticlastic and non-aceticlastic Methanosarcinales, with exception of Methanosarcina species. This distribution supports the scenario where the complete loss of ACS and PPase occurred only in Methanosarcina (after the acquisition of the ackA/pta system), whereas the particular loss of the pyrophosphatase PPase occurred only in strict methylotrophic Methanosarcinales. Second, the presence of both enzymatic systems to activate acetate is observed so far in one single aceticlastic species within Methanosarcina, *Methanosarcina siciliae* strain C2J, suggesting that the ACS and PPase was indeed present in Methanosarcina before. Third, there is indeed a difference in the enzymatic efficiency of both systems based on in their affinity to acetate. Under high concentrations of acetate, for instance, the ACS/PPase system shows a low activity but a high affinity for it, whereas the ackA/pta system shows a high activity but a low affinity for acetate (Min and Zinder 1989). This suggests than an enzymatic shift from an inefficient to an efficient system to activate acetate at high concentrations in Methanosarcina might have contributed to outcompete Methanosaeta species in such conditions (Jetten et al. 1992).
Finally, the earlier and later phylogenetic origins of Methanosarcina and Methanosaeta within Methanosarcinales, respectively, as estimated by (Rothman et al. 2014) (supplementary figure SF10), also contributes to infer the independent acquisition of these different systems to activate acetate in Methanosarcinales. Nevertheless, the findings of this thesis are in partial agreement with the study of Rothman et al., (2014). These authors have suggested that basal species of the Methanosarcina group, such as *M. baltica, M. semesiae and M. lacustris,* do not perform aceticlastic methanogenesis because they do not grow under aceticlastic conditions and do not possess the ackA/pta system. And thus, such

methanogenic pathway should have appeared in recent species of Methanosarcina. However, I found that *M. baltica* grows indeed under aceticlastic conditions according to literature (Singh et al. 2005; von Klein et al. 2002), and that *M. lacustris* contains the ackA/pta system according to my genomic findings. Yet, the identification of the ackA/pta system in *M. baltica* will remain unclear until its complete genome is available, whereas the incapability of *M. lacustris* to grow under aceticlastic conditions (Simankova et al. 2001), (despite of containing the ackA/pta system) could be explained by the loss of the Frh hydrogenase. The Frh complex has been proved to be crucial during the aceticlastic methanogenesis, and more efficient that its functional counterparts Fpo and Vht, during the reduction of heterodisulfide methyl-coenzyme M (CoM-SH) and coenzyme B (CoM-S-S-CoB) in freshwater species (Kulkarni et al. 2009).

On the other hand, it is worthy to also note the HGT event that was identified from archaea to bacteria involving the enzyme that cleavage the acetyl-CoA after the acetate activation, i.e. archaea-type carbon monoxide dehydrogenase/acetyl-CoA decarbonylase complex (CODH/ACDS). A manual inspection of the gene trees suggests that this transfer might have occurred from a methanogen (likely from Methanosarcinales) to the single recipient bacterium *Desulforudis audaxviator,* as previously suggested (Chivian et al. 2008; Techtmann et al. 2012). The presence of no homologs of the archaea-type CODH/ACDS (cdhABC) in other bacteria support this single export transfer event from archaea to *Desulforudis audaxviator*.

Methylotrophic methanogenesis. The presence of bacterial HGTs in specific genes of this methanogenic pathway has been reported only for *M. acetivorans* (Deppenmeier et al., 2002). In this study, I found that most of the genes involved in the first steps of the methylotrophic pathway in Methanosarcinales are the evolutionary outcome of HGT events most likely from bacteria. From the 12 methyltransferases (MTs) associated to this pathway, my findings show that 11 present HGTs (5 classified as imports), while 1 has no bacterial homolog within the protein family. Furthermore, gene copies of the MTs, such as those using methanol (mta) and monomethylamine (mmt), only have distant bacterial homologs, but no close homologous counterparts within their corresponding gene families. Indeed, the distribution of the MTs is restricted across prokaryotes, in agreement with previous reports (Borrel et al. 2014; Prat et al. 2012; Rother and Krzycki 2010). In Methanosarcinales, for example, MTs related to dimethylamines (DMA) and dimethylsulfide (DMS) are absent in some obligated methylotrophic methanogens and marine species, respectively. An explanation of such absence is that some members of Methanosarcinales might use a different methyltransferase for DMA or DMS that have not been identified so far. Another scenario involves the disruption of the enzymatic function of some MTs. For instance, it has been found that one MT (*mtbB*) in the obligated methylotrophic methanogen *M. burtonii* has been disrupted by a putative transposon, leading to its annotation as a pseudogene in KEGG and further suggesting a loss of function (Allen et al. 2009). Similar effects might be expected for other genes in Methanosarcinales and other archaea, given that they have a considerable number of transposases and other insertion sequences in their genomes (Brügger et al. 2002; Filée et al. 2007; Maeder et al. 2006).
On the other hand, the gene trees used to perform the "density distribution of species" approach cannot offer a credible evidence to classify some of the transferred MTs as imports or exports. As shown in Figure 5.2 (in Methods), such uncertainty is because it is not possible to define a clear direction of a particular transfer when the number of bacterial and archaeal clades is equivalent or similar and with a restricted distribution across prokaryotes. The limitation of this approach has been previously described by Koonin et al. (2001), who has suggested careful inference of these cases. Thus, any designation as imports or exports were omitted, specifically for methylated amines (*mmtBC*, *mtbABC*, *mttBC*), methanol (*mtaABC*), and methyl sulfides (*mtsAB*), as well as for those

genes (*pylSBCD*) related to the pyrrolysine biosynthesis. On the contrary, the gene trees related to all permeases involved in the uptake of methylotrophic substrates (mtmP, mtbP, mttP), were manually verified and classified as imports, representing reliable cases of bacterial acquisitions by Methanosarcinales. Collectively, these findings cannot support a complete bacterial origin of this pathway via HGT in Methanosarcinales. Further analyses integrating recently discovered methylotrophic methanogens in archaea, such as *M. stadtmanae* (Miller and Woling 1985), Methanomassiliicoccus (Borrel et al. 2014), Bathyarchaeota (Evans et al. 2015), 'Verstraetearchaeota' (Vanwonterghem et al. 2016), and 'Methanonatronarchaeia' (Sorokin et al. 2017), will help to elucidate the evolution of this pathway in archaea and to discriminate between donor and recipient in the controversial cases of HGT.

7.10.4. On the origin and lost of the hydrogenotrophic pathway in Methanosarcinales.

The hydrogenotrophic methanogenesis is considered as the most ancient and widespread pathway in Euryarchaeota, and is indeed conserved in almost all methanogens (Thauer et al., 2008). Nevertheless, it is also well known that some members of Methanosarcinales are incapable to grow in hydrogenotrophic conditions, such as the obligated aceticlastic and methylotrophic methanogens (.a.k.a methanogens substrate-specific), or the facultative *Methanosarcina acetivorans* (Sowers et al., 1984). As detailed in the Result section and Figure 7.13, from the 12 methanogenic species (out the 16 compiled with experimental reports) that were tested to growth under hydrogenotrophic conditions, 7 species were not able (and 2 additional species showed difficulties) to grow under such conditions. There are several reasons that could explain why some species are apparently not able to perform hydrogenotrophic methanogenesis: 1) changes in the *culture medium* (Maestrojuán and Boone 1991); 2) limitation in the *amount and availability of resources in the environment* to perform this methanogenesis, such as the H_2 concentration and its effect on gene regulation and growth rate in methanogens (Hendrickson et al., 2007); and 3) the *genomic variation* of some genes in this pathway that (Allen et al. 2009; Walter et al. 2016) (Allen et al., 2009; Walter et al., 2016) might derive in alternative phenotypes (Hunt et al. 2011). For instance, certain phenotypes can produce methane in limited or unusual amounts (Hutten et al. 1980; Mah 1980; Peinemann et al. 1988; Shi et al. 2014), whereas different strains from the same species are not necessarily able to use the same substrate to perform hydrogenotrophic methanogenesis, such as the strains S6 from *M. mazei* (Maestrojuán and Boone 1991), *M. acetivorans*, *M. baltica*, *M. siciliae* T4M, and others (Sowers et al. 1984; von Klein et al. 2002). Although several of the Methanosarcina species reported in the literature have not been experimentally tested to perform methanogenesis with all known substrates, the previous evidence clearly suggests that several Methanosarcinales species have loss the ability to perform hydrogenotrophic methanogenesis.

The genomic evidence suggesting the loss of the hydrogenotrophic pathway has been already suggested in multiple clades of archaea (Bapteste et al. 2005). This observation can be also extended to members of Methanosarcinales, given the experimental observations and the genomic evidence estimated in this work. To start with, the two key H_2-dependent components, the cytoplasmic coenzyme F_{420}-reducing hydrogenase (Frh) and the membrane energy conserving hydrogenase Ech, are found to be conserved only in species of the Methanosarcina genus living in freshwater environments. By contrast, the Ech and Frh complexes are found to be lost in Methanosarcinales species living either in marine environments or wastewater treatment reactors. Notably, these groups of marine and wastewater species have different characteristics, for example, the marine species belonging to the *Methanosarcinaceae* family have large genome sizes and are present in syntrophic interactions (Worm et al. 2014), others species from the *Methanosaetaceae* family have very small

genomes and represent mostly endosymbionts (Narayanan et al. 2009), while another species live in marine environments (Mori et al. 2011). Following this line of though, previous reports have also underscored the energy-based differences between H_2-dependent (freshwater) and the H_2-independent (marine) Methanosarcinales (Allen et al. 2009; Ferry and Lessner 2008; Li et al. 2006). For instance, the thermodynamic constrain in the H_2 concentration threshold needed by methanogens to perform hydrogenotrophic methanogenesis (Thauer et al. 2008) might have led to other H_2-dependent organisms to use alternative substrates in a competitive environment (de Poorter et al. 2007), such as formate by hydrogenotrophic methanogens without cytochromes (Wood et al. 2003). Other studies have suggested that H_2-dependent Methanosarcinales could be outcompeted by other strict H_2-dependent organisms, such as sulfate reducing bacteria (Conrad 1999). Thus, the absence of the hydrogenases Frh and Ech in saltwater Methanosarcinales might explain, in part, the incapability of obligated methylotrophic methanogens to grow in hydrogenotrophic conditions (i.e., no reduction de CO_2). Other scenarios have to be considered for the substrate-versatile in Methanosarcina species. For instance, it seems that the Rnf and Fpo complexes have replaced the functions of Ech (Ferredoxin reduction) (Li et al. 2006; Welte and Deppenmeier 2013) and of Fpo (oxidation of $F_{420}H_2$) in methylotrophic methanogens, respectively (Greening et al. 2016; Guss et al. 2009; Ney et al. 2017). Surprisingly, the growth of Methanosarcina species in methylotrophic conditions suggests that Fpo appears to be dispensable, while the Frh seems to be essential (Ney et al. 2017; Welte and Deppenmeier 2013).

On the other hand, the five genes involved in the biosynthesis of the cofactor F_{420} (*cofC*, *cofD*, *cofE*, *cofG*, *cofH*) were found to be part of HGT events across Methanosarcinales, which is in agreement with the findings of a recent study (Ney et al. 2017). The cofactor F_{420} is not only used in the hydrogenotrophic (involved in CO2 reduction) and methylotrophic (respiration process) methanogenesis (Greening et al. 2016), but it also involved a wide range of redox reactions from multiple biological processes in archaea and bacteria. Accordingly, my analysis reflects a wide distribution of these genes across Euryarchaeota without a clear phylogenetic pattern. Furthermore, there are difficulties to clarify the origin of the HGT events involved in the cofactor F_{420} biosynthesis, which suggest the presence of at least two independent evolutionary events. Consistent with the Ney et al. study (2017), the analysis of the phylogenetic gene trees involved in the biosynthesis of F_{420} (*cofC*, *cofD*, *cofE*) found in archaea (Thaumarchaeota, Archaeoglobales, Halobacteriales, all methanogenic clades) and bacteria (Actinobacteria and Proteobacteria, Chloroflexi, Thermatogae, and Spirochaetes) indicates that the transfer of these genes were imported from bacteria to archaea. In contrast, the analysis of the phylogenetic gene trees involved in the biosynthesis of F_o (*cofG*, *cofH*)—only found in archaea and specific members of Cyanobacteria and Proteobacteria—indicates that these genes were most likely exported from archaea to bacteria, being an early actinobacterium the potential ancestral recipient, according to Ney et al. (2017). Nonetheless, additional studies are necessary to clarify the origin and evolution of these genes. Altogether, these findings support the idea that the genes *cofC*, *cofD*, *cofE* where transferred to archaea after the origin of *cofG*, *cofH* in archaea, given that the synthesis of F_{420} depends on the biosynthesis of F_o. Although the transfer of *cofG* and *cofH* from archaea to bacteria is not conclusive in the present study, it supports the hypothesis of Ney et al. (2017) that these potential exports occurred earlier in the origin of archaea.

Lastly, additional HGTs were detected in gene copies of enzymes playing a role in key reactions of this pathway, such as *mtrA*, fdhA, and *fwdA*, *fwdB*, *fwdC*, *fwdD*, and *fwdG*. The manual inspection of the non-monophyletic trees of these genes suggests that these transfers occurred from archaea to bacteria (exports). The distribution of these genes in bacteria corresponds mainly to Proteobacteria, while Clostridia and Planctomycetia are represented in a minor amount. Particularly, bacterial

homologs of these genes have been found in methanotrophs (Karrasch et al. 1989; Pomper et al. 2002; Pomper et al. 1999; Vorholt et al. 1999), and methylotrophs (Chistoserdova et al. 2005; Chistoserdova et al. 1998; Pomper et al. 1999; Vorholt et al. 1999), but they are also found widespread across prokaryotes (Kalyuzhnaya and Chistoserdova 2005; Kalyuzhnaya et al. 2005; Kalyuzhnaya et al. 2004). In agreement with Galagan et al. (2002), this suggests that these genes have been very likely subject to several HGTs events across vast phylogenetic distances.

In perspective, the production of methane is the only energy-yielded metabolism in Methanosarcinales, and is carried out by three main and different pathways, having only in common the demethylation of methyl-coenzyme M to methane and the reduction of the heterodisulfide of coenzyme M and coenzyme B catalyzed by methyl coenzyme M and heterodisulfide reductases (Ferry 2002; Ferry 2011; Thauer 1998; Thauer et al. 2008). Collectively, the findings of this thesis suggest that considerable genomic changes, including key gene losses and horizontal gene gains from bacteria, have contributed to the origin and evolution of methanogenesis in Methanosarcinales. First, in the genomic changes behind the energetic-based diversification of the methanogenic pathways, thanks to which some species have been able to retain a high energy production under hydrogenotrophic ($DG^{o\prime}$=–131 kJ/mol^{-1}) and methylotrophic ($DG^{o\prime}$= from –78.7 to –191.1 kJ/mol^{-1}) conditions (Oren, 1999), whereas aceticlastic species maintain their energy production under the limits ($DG^{o\prime}$= –36 kJ/mol^{-1}) (Welte and Deppenmeier et al., 2013). And second, in the ecological transition from freshwater and saltwater environments through the lateral acquisition of particular enzymes that allowed the use of noncompetitive substrates, such as the methylated compounds.

8

THE BACTERIAL ACQUISITION OF THE AEROBIC RESPIRATION PROTEIN COMPLEXES IN HALOBACTERIALES HAS REMAINED WELL CONSERVED AND IS ORGANIZED IN AN OPERON-LIKE MANNER

Halobacteriales (Haloarchaea) represent one of the most extreme groups of microorganisms that are capable to live in a wide range of hypersaline environments. They are also obligated heterotrophs and are either obligated or facultative aerobes. The aerobic respiratory chain system has been well studied in prokaryotes and consists of five membrane-bound protein complexes, each one represented in a single or two versions, depending of the gene(s) configuration of the corresponding complex that is used by the organism. For instance, complex I is formed by the *NADH dehydrogenase* or the *NAD(P)H dehydrogenase*, complex II is represented by the *Succinate dehydrogenase* or the *Fumarate reductase,* complex III is formed by the *cytochrome* bc_1, complex IV is represented by the *cytochrome c oxidase* or the *cytochrome bc oxidase*, and complex V has only one version formed by the *A-type ATPase*. In a previous study that analyzed the impact of HGT events in 10 genomes of Halobacteriales (Nelson-Sathi et al. 2012), Martin and coauthors suggest that this archaeal clade appears to be the result of the lateral acquisition of hundreds genes from Bacteria. In particular, this work underlines the bacterial acquisition of the aerobic respiratory system, suggesting 213 gene candidates as part of the complexes I to IV. Overall, the authors suggested that such massive HGT led to the transformation from an *anaerobic autotroph* lifestyle to an *aerobic heterotroph* lifestyle.

In the present research, I re-assessed the presence and characterization of HGT events across all genes from the five membrane-bound protein complexes that form the aerobic respiratory chain system of 23 Halobacteriales genomes. A further goal was to take into account the synteny conservation of such genes, that is the physical co-localization of genetic loci on the same chromosome, at two genomic organization levels: gene operon-like structure within each membrane respiratory complex and the operon-clustering of all five membrane-bound protein complexes along a genome. It is well known that the bacterial counterparts of the five complexes are organized in operons (Guest et al. 2017; Price and Driessen 2010; Richardson 2000). However, it has been reported that Halobacteriales can harbor multiple gene copies in either close genomic proximity or distantly located from each other along the genome (Bueno et al. 2012; Zerulla and Soppa 2014). It is interesting thus to understand whether the bacterial acquisition of the aerobic respiration chain at the ancestor of Haloarchaea has been also

followed by synteny conservation.

To that end, a manual curation of the genes involved in the aerobic respiration chain was performed first (see Methods). Briefly, 213 KO numbers (*a.k.a* K identifier), involved in the oxidative phosphorylation process that encompass the five protein complexes, were obtained from the KEGG gene orthology annotation (see Supplementary figure SF11), and they do not correspond to the 213 genes reported by Nelson-Sathi et al. (2012). Accordingly, these identifiers represent all known annotations related to Archaea, Bacteria and Eukarya, and their distribution across the 1,981 prokaryotic genomes used along this study is shown in the Supplementary figure SF12, Next, only those K identifiers present in Archaea and/or Bacteria, or common in all domains (i.e., cytochrome c reductase) were retained for further analyses, whereas K identifiers belonging exclusively to Eukarya were removed. To identify HGT events, the gene candidates from the five membrane-bound protein complexes in Halobacteriales were mapped across the previously created 25,762 archaeal gene families (Chapter 6) to identify the homologs between Halobacteriales and bacteria. Through the same methods described in methods, HGT events were then classified as imports, exports and "undefined" transfers. The synteny conservation was further analyzed, and finally a comparison between the previous findings by Nelson-Sathi et al. (2012) and the present work is assessed in this chapter.

8.1 The genes representing the aerobic respiratory complexes I, II, III, IV, and V are well conserved across Halobacteria.

The figure 8.1 shows the gene distribution of orthologs related to the respiratory chain system in Halobacteriales, as defined by their K identifier names. After a manual inspection, it was clear to recognize that the absence of multiple subunits across the five protein complexes is due to the redundancy of identifying a gene with different names. In the complex III, for instance, the subunits *qrcB*, *cytB* and *petB* are the same and correspond to the cytochrome b. Therefore, all genes named as one of these cases were collapsed as one representative gene of the cytochrome b. The same approach was performed for the subunits named as *qrcC*, *cytC* and *petC,* which all correspond to cytochrome C, as well as the gene subunits named as *qrcA*, and *petA* that represent the Rieske iron sulfur subunit from complex III (see Figure 8.1). Thus, only the genes forming the complete versions of each membrane complex or on each Halobacteriales genomes were considered in this study.

Accordingly, a final set of 1,179 genes representing 44 K identifiers (from the KEGG database) was used to depict the aerobic respiration system in 23 Halobacteriales genomes. The previous genes and K identifiers are distributed along the membrane-bound complexes as follow: **complex I** is defined by the NADH dehydrogenase (*nuoA-N*); **complex II** by the *succinate dehydrogenase* (*sdhABCD*); **complex III** is defined by the Rieske iron sulfur protein (ISP), *b*-type cytochromes (CytB), and *c*-type cytochrome proteins (CytC); **complex IV** is defined in two versions, one by the *cytochrome c oxidase (coxABCD)* and the other by the cytochrome bc oxidase (*cydABC*); and **complex V** is defined by the A-type ATPase (*atpA-H)*. From this large set of homologs genes, the next step was identify orthologs that form clusters of genes organized in an operon-like manner and, therefore, excludes gene copies from further analyses.

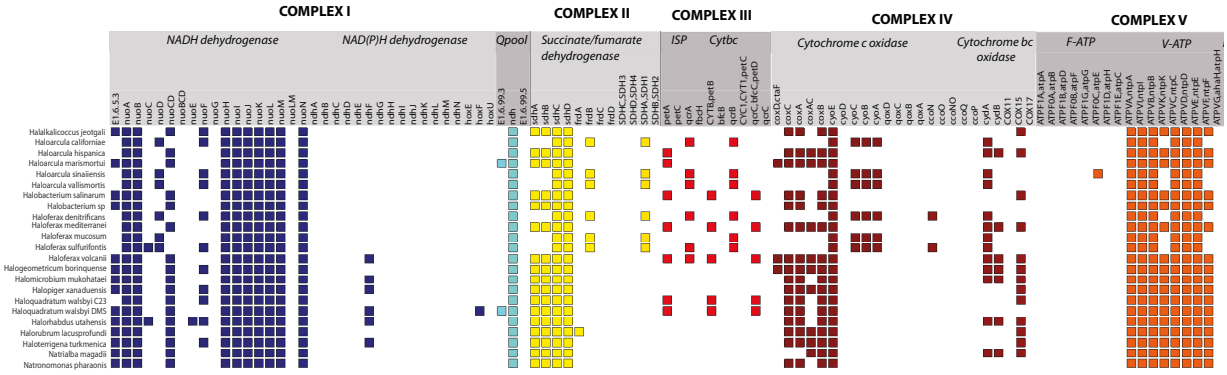


Figure 8.1. Distribution of genes containing K identifiers from KEGG involved in the aerobic respiratory chain system in Halobacteriales. Distribution of the genes (squares), designated by K identifiers, across each membrane-bound protein complex (at top) in the 23 Halobacteriales genomes (left axis). The names corresponding to the five membrane-bound protein complexes, as well as for the genes are shown at top. Halobacteriales genomes are sorted phylogenetically based on the reference tree obtained from universal proteins (described in Chapter 6). For each complex there are more than two version of the set of proteins. For example, the dehydrogenases *NADH* and *NADPH* correspond to Complex I, while Complex II is represented—in some cases—by either *succinate* or *fumarate dehydrogenase*. Also the Complex IV is presented either by *cytochrome C* or *cytochrome bc oxidase*, and the Complex V by a bacterial *F-type* or archaeal *A-type* ATPase synthase.

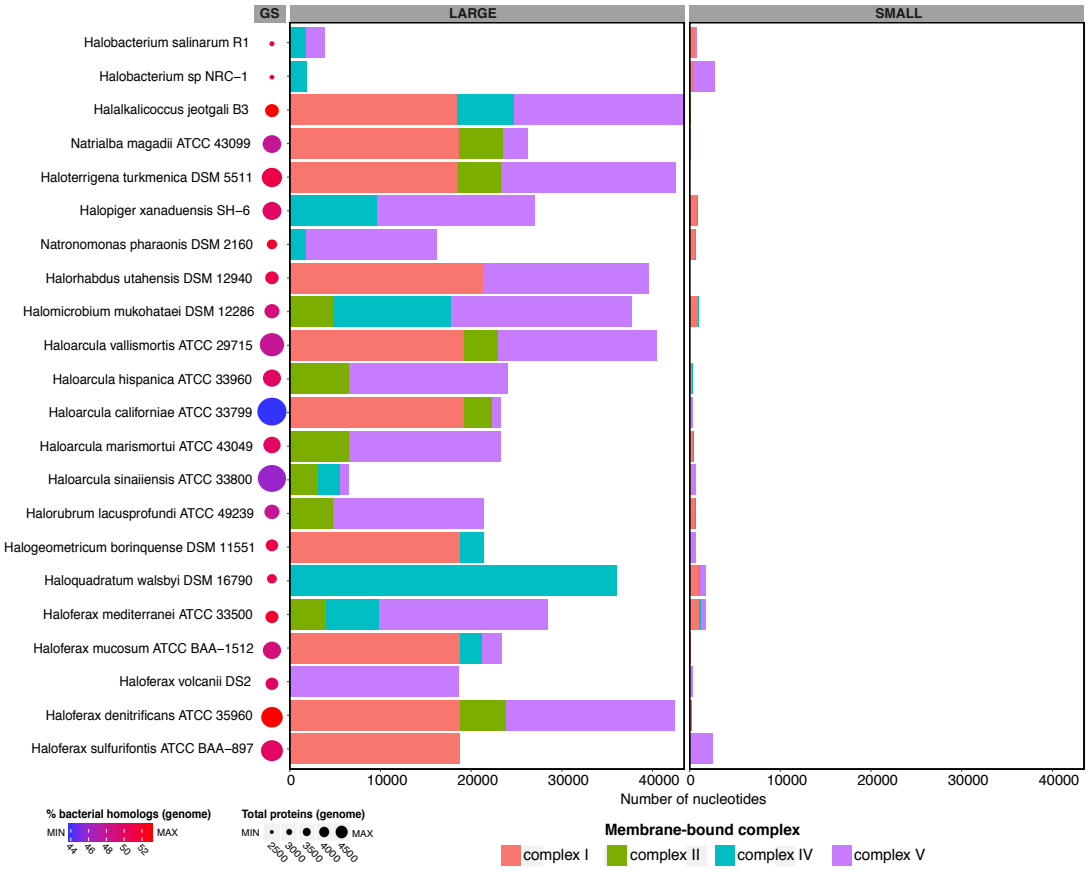


Figure 8.2. Distribution of large and small intergenic distances of each membrane complex of the aerobic respiratory chain system in Halobacteriales. Large intergenic (at left) and small (at right) spaces between pairs of genes are shown with the same scale for each membrane complex in all 23 Halobacteriales genomes sorted phylogenetically (left axis). Largest distances occur in Complex I, IV, and V. Proteome size of Halobacteriales (circle size) and the percentage of bacterial-like genes contained in their proteomes (scaled bar at bottom-left side). Membrane complexes are classififed by colors (see squared colors at bottom).

8.2 The operon-like organization of genes forming each respiratory complex is conserved across Halobacteriales, but they synteny of the five operons is not.

Essentially, the identification of an operon-like organization on each of the membrane-bound complexes was performed under four criteria described briefly as follow (see also Methods): **1)** that **adjacent gene pairs** belonging to a specific membrane bound complex are located in the same strand, **2)** that a **minimum number of subunits** composing a specific protein complex is conserved across genomes (see Table 5.6 and Methods), **3)** the **insertion or deletion of genes between conserved subunits** is allowed, given that one or more genes can modify pre-existing operons (Price et al. 2006), and 4) only the most complete copy of the whole operon-like organization is taken into account. It is important to note that an **intergenic distance threshold** (I_{dT}) was not used as a criterion to identify synteny conservation because this can vary widely across species and genome sizes (Nuñez et al. 2013). However, an *ad hoc* estimation of I_{dT} was calculated to assess the intergenic distance between two genes within the same operon-like structure. As observed in Figure 8.2, small intergenic distances were measured as those between ≥10 and ≤300 nucleotides (nts), whereas large intergenic distances were estimated as those of ≥1,000 nts.

Through this approaches, a search of all prokaryotic homologs from the 1,179 genes related to the protein complexes was carried out through the 25,762 archaeal protein families characterized in Chapter 6. From this searching procedure, homologs were found in 88 protein families and the operon-like organization of these genes and for each of the five membrane-bound complexes across all 23 Halobacteriales genomes was reconstructed with the rules previously described (see panel B in Figure 8.2). From this analysis, 675 genes represent as complete operon-like structures (53%), while the remaining 504 genes (43%) represent individual gene that do not cluster with any other or forming incomplete operon-like clusters (see "operon detection" section on panel A of Figure 8.3). The 504 genes were designed as *gene copies*. Interestingly, the number of gene copies forming the five membrane-bound complexes in a genome seems to be related to its proteome size (see circle sizes in the "Proteome size" on panel A of Figure 8.3). Comprehensive plots illustrating the distribution of operon-like structures and non- or incomplete operon-like structures belonging to each protein complex across all Halobacteriales genomes are provided in supplementary figures SF13-SF17.

The total amount of genes corresponding to each protein complex, as well as their operon-like organization are conserved across Halobacteriales (see subunits and genomic organization of operons in *panel B* of Figure 8.3). For example, **complex I** in all Halobacteriales is composed by 11 subunits: *nuoA, nuoB, nuoC, nuoD, nuoH, nuoI, nuoJ nuoK, nuoL, nuoM, nuoN;* although in some cases, *nuoJ1-nuoJ2* are merged, and the bacterial module nuoEFG (*nuoE*, *nuoF* and *nuoG*) was found in some species, but they are not associated to the cluster of genes defining the operon. The **complex II** in all Halobacteriales is defined by 4 subunits (*sdhA*, *sdhB*, *sdhC*, and *sdhD*), and the **complex III** is formed by three subunits (CytC, CytB, and the Rieske protein, usually known as *petA*, *petB* and *petC*, respectively). Likewise, the **complex IV** is represented by two versions across Halobacteriales, one version is formed by the three subunits *coxA, coxB* and *coxC*, and the second version encompass *cydA* and *cydB* (a.k.a *cydAB*), which is present only in some species. Finally, the **complex V** is composed with 9 subunits (*atpD*, *atpB*, *atpA*, *atpF*, *atpC*, *atpE*, *atpK*, *atpI*, and *atpH*) across all Halobacteriales. Nevetheless, some subunits of specific complexes are not found missing; for instance, sdhC in complex II (*Haloferax meditarrani*), *cytB* and *cytC* in the complex III of four species (*Natrialba magadii*, *Haloterrigena turkmenica*, *Halopiger xanaduensis*, and *Natromonas pharaonis*), and *coxD* in the complex IV of several species.

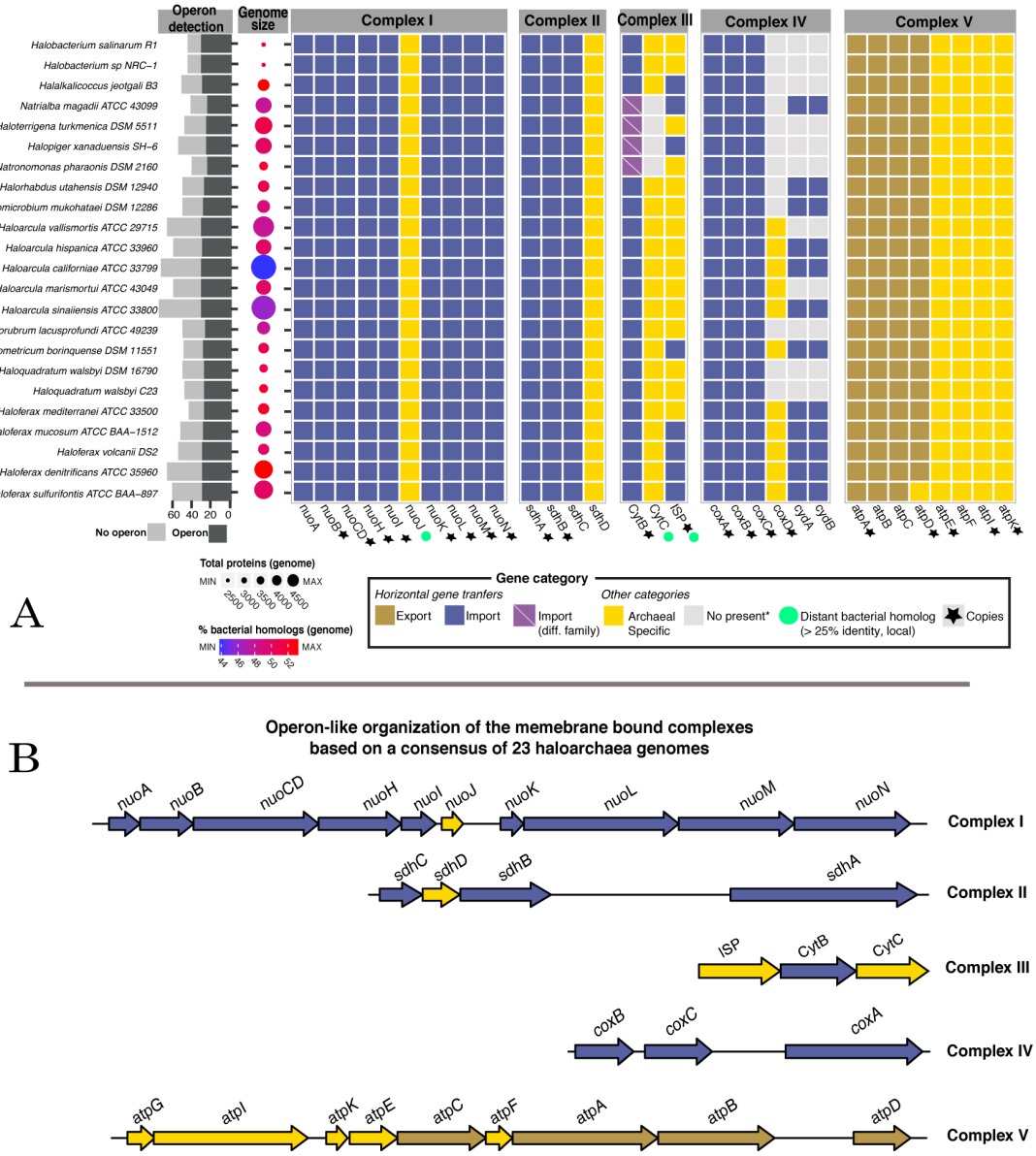


Figure 8.3. Distribution of genes in each membrane-bound complexes involved in the aerobic respiration within Halobacteriales genomes and the corresponding operon-like organization model. Conservation of gene subunits in for each complex across Halobacteriales genomes, and their HGT classification (upper figure A). Schematic operon-like structure models are shown for the five membrane complexes (lower figure B). **Upper Figure A:** Halobacteriales genomes are sorted phylogenetically (as described in Chapter 6). Next, the total number of genes identified as part of the operon-like organization (operon) or gene copies (No operon) is shown for each species (see bar plot). The proteome sizes for Halobacteriales genomes are scaled circle sized (see "total proteome size") and colored scaled (see "% bacterial homologs"). Presence or absence of HGTs in genes of the complexes in Halobacteriales were identified, and defined by colors at the bottom (see "Gene category"). Distant homologs were identified (≥25 identity locally aligned). The absence of genes (*, No present) is defined either by a complete lack of the gene in the genome or because the method used here could not recovered such gene. **Lower Figure B:** Schematic operon-like organization was obtained for all membrane-bound complexes based on a consensus of 23 haloarchaea genomes. The operon-like structure (number and order of genes) does not vary across all members of Halobacteriales (see Figure 8.4). In each operon, the arrows represent genes and the colors in it represent either the types of HGTs or the absence of HGT (see 'Gene category' at upper panel). Lines connecting genes represent intergenic distances; the corresponding distances (in kilobases) and its variation within and across each halobacteriales is shown in the Figure 8.2 and Figure 8.3, respectively. Note that large distances are present once (specific location), and not repeatedly across the entire operon.

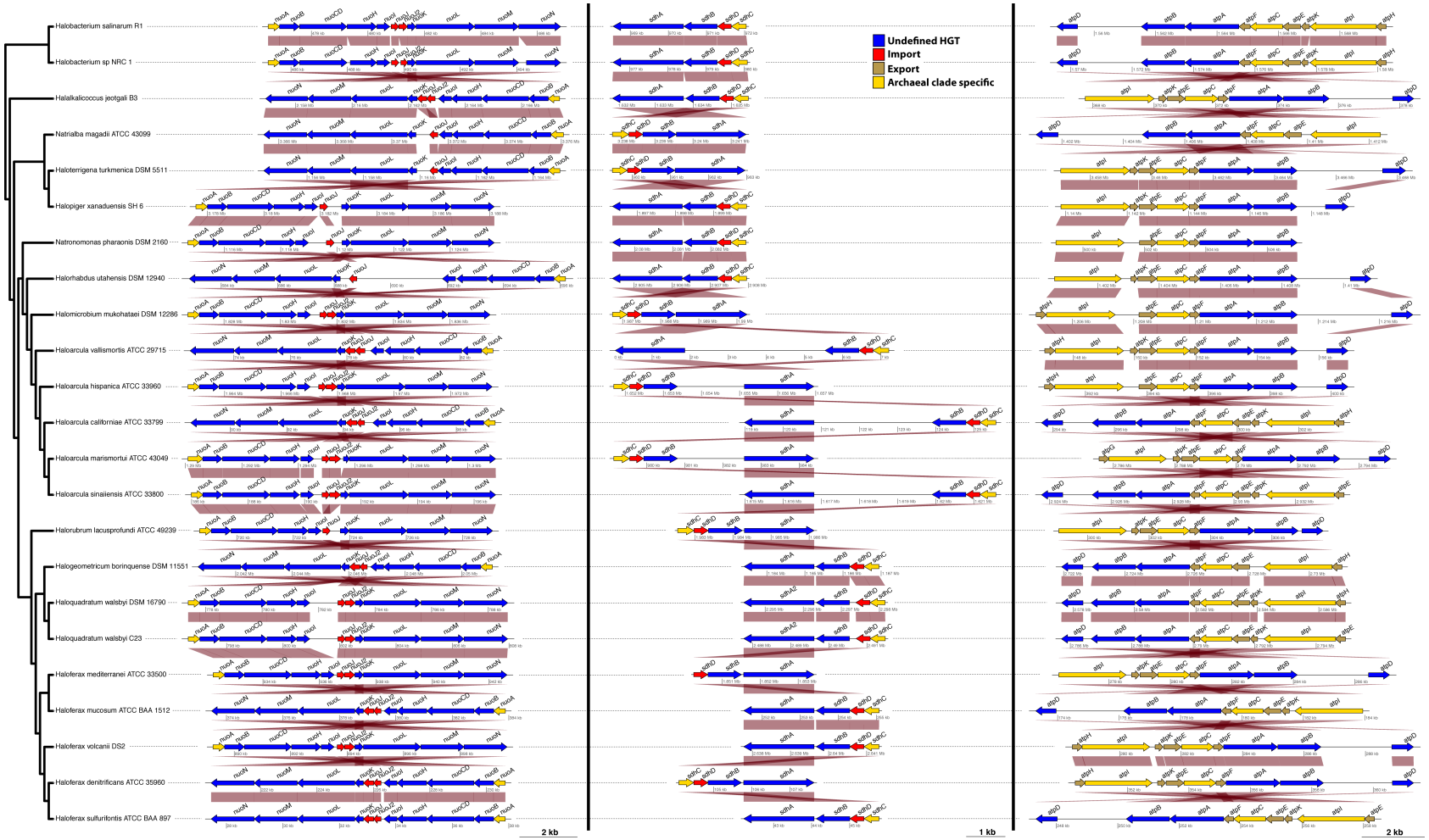


Figure 8.4. Operon-like organization structures for complexes I, II and V within and across the 23 Halobacteriales species. Halobacteriales species are ordered phylogenetically according to the phylogenetic reference tree reconstructed in Chapter 6 (*left side*). Cluster (operon-like) of genes were found either in the negative (arrows to the right direction) or positive (arrows to the left direction) strand in each genome. The presence or absence of HGT is represented by colors as the classification of HGTs recovered from each tree topology, i.e. import (red), export (gold), and undefined transfer (blue), or as archaeal-specific (yellow). Gene names for each subunit are shown at the top of each arrow. The coordinates of the genes are scaled in kilobase pairs (kb).

According to the above, the subunits from I to V of the membrane-bound protein complexes are well conserved across the 23 Halobacteriales. However, the exception to this trend is represented by the two versions of complex IV, *coxABCD* and *cydABC,* as a probable result of one complex used in the absence of the other one. For instance, *Halobacterium sp.* strain NCR-1 contains three different cytochrome oxidases (*cbaABD*, *coxABC*, and *cydAB*) that are involved in aerobic respiration (Müller and DasSarma 2005). As observed in Figure 8.3, *coxABC*, and *cydAB,* are both present in almost half of the species in Halobacteriales, except for *H. californiae, H. sinaiiensi, H. vallismortis, H. mucosum,* and *H. sulforifontis*, were only one version of the complex IV is very likely used in the respiration process.

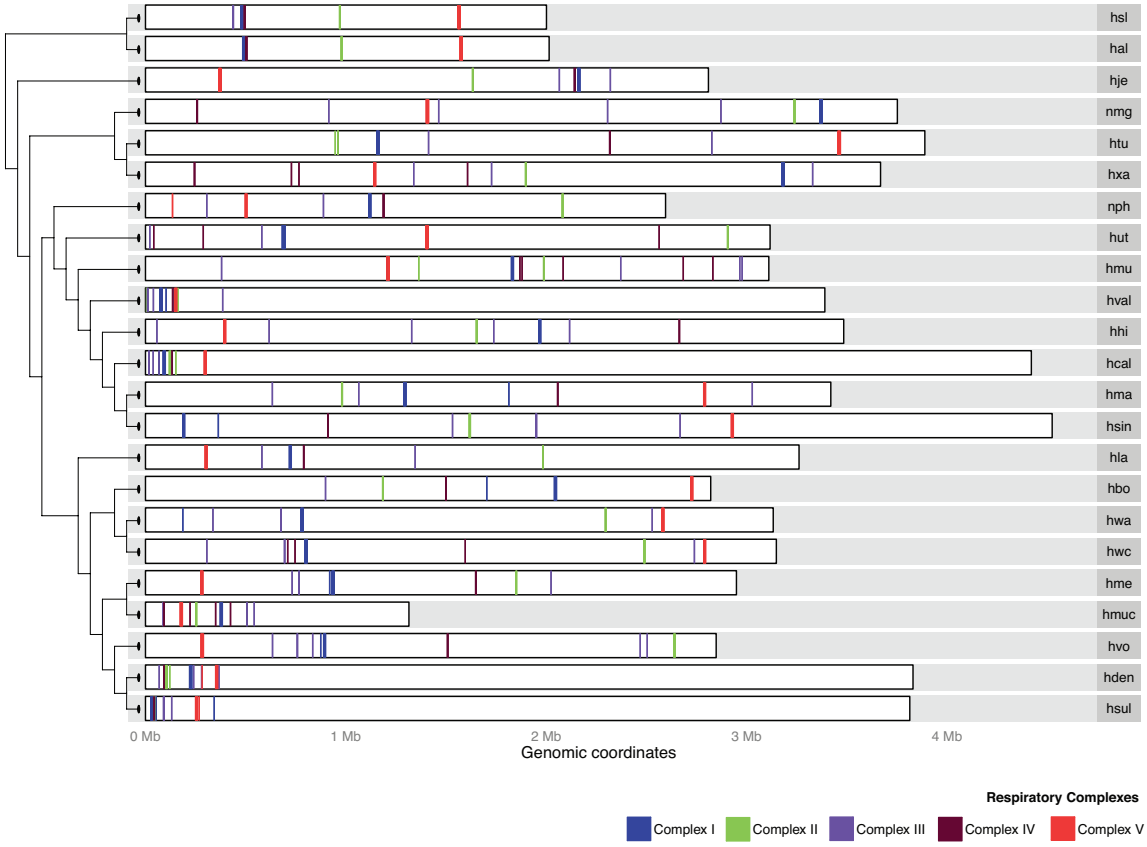


Figure 8.5. Whole synteny conservation of the 5 membrane-bound complexes within the Halobacteriales genomes. Halobacteriales species are ordered phylogenetically according to the phylogenetic reference tree reconstructed in Chapter 6 (*left side*). The location in the genome of each complex is shown according to the genomic coordinates quantified in megabases (lower axis), and are distinguished by colors (see color squares). Gene copies previously reported in this study were not included in this genomic map. The corresponding names of Halobacteriales members are displayed in code letters format (see grey colored squares at right side). Complete species names of Halobacteriales are provided from the top to the bottom: *Halobacterium salinarum* R1 (hsl), *Halobacterium sp.* NRC1 (hal), *Halalkalicoccus jeotgali* B3 (hje), *Natrialba megadii* ATCC 43099 (nmg), *Haloterrigena turkmenica* DSM 5511, *Halopiger xanaduensis* SH-6 (hxa), *Natronomonas pharaonis* DSM 2160 (nph), *Halorhabdus utahensis* DSM 12940 (hut), *Halomicrobium mukohataei* DSM 12286 (hmu), *Haloarcula vallismortis* ATCC 29715 (hval), *Haloarcula hispanica* ATCC 33960 (hhi), *Haloarcula californiae* ATCC 33799 (hcal), *Haloarcula marismortui* ATCC 43049 (hma), *Haloarcula sinaiiensis* ATCC 33800 (hsin), *Halorubrum lacusprofundi* ATCC 49239 (hla), *Halogeometricum borinquense* DSM 11551 (hbo), *Haloquadratum walsbyi* DSM 16790 (hwa), *Haloquadratum walsbyi* C23 (hwc), *Haloferax mediterranei* ATCC 33500 (hme), *Haloferax mucosum* ATCC BAA-1512 (hmuc), *Haloferax volcanii* DS2 (hvo), *Haloferax denitrificans* ATCC 35960 (hden), *Haloferax sulfurifontis* ATCC BAA-897 (hsul).

Although the intergenic distances (I_d) among genes within each complex vary widely between Halobacteriales species, the most variable and large intergenic distances in are observed in the complexes I and V. As also observed in Figure 8.4, the largest distances occur mainly in the complex I, complex II, and complex V, and specifically between the subunits *nuoJ-nuoK, sdhB-sdhA*, and *atpB-atpD* and *atpI-atpE*, respectively. These cases occur in specific genera, such as Halalkalicoccus, Natrialba, Haloterrigena, Haloarcula, Haloferax, and Halorhabdus.

The conservation of the five operons representing the aerobic respiratory complexes along a genome was also performed, based on manual inspection, and by mapping the genomic coordinates of the five membrane complex operons to create creating individual maps for each genome in Halobacteriales. As observed in Figure 8.5, the distribution of the five operons along the Halobacteriales genomes, at least the complete version identified, is highly widespread, not conserved and none uniform patterns, rather than in those strains from the same species (e.g. *Halobacterium salinarum* and *Halobacterium sp.*) and the same genus (e.g. *Haloarcula vallismortis* and *Haloarcula californiae*).

8.3. The influence of HGT events in the evolution of the aerobic respiratory system in Halobacteriales.

As mentioned previously, a total of 675 genes represent the set of genes that are clustered as operon-like of the five membrane protein complexes in Halobacteriales. These genes were found within 50 protein families, where 20 are identified as archaeal specific and 30 have bacterial homologs (see Table 8.1). The phylogenetic reconstruction of these families and the analysis of the direction of the HGTs (as described in Methods), allow to identify the HGTs, and classify these cases as follow: 9 gene families were classified as imports, 5 as exports, and the remaining 16 as undefined transfers (see headers names of columns colored in grey of Table 8.1). For these undefined transfers, the direction of the transfer was attested by performing manual inspections of each phylogenetic tree. In all of the cases, Halobacteriales proteins were branched together (monophyly), however, additional archaeal proteins from other clades were widely distributed in the tree interleaving among bacterial branches (disruption that classify them as non-monophyletic). Nevertheless, it confirmed that these genes were indeed bacterial acquisitions, and as observed in Figure 8.2, the membrane-bound protein complexes I, II, III, and IV of the aerobic respiration system in Halobacteriales were acquired from Bacteria via HGT. However, the genes *nouJ, sdhD,* and *coxD* were designed as archaeal specific in complexes I, II, III, and IV, respectively (see panel A in Figure 8.2). Interestingly, the gene families of these cases do not have bacterial homologs (a ≤25% identity); however, they have distant bacterial homologous gene with a low ≤30 similarity percentage (local alignment) (see circles colored cyan at bottom in Figure 8.2). In contrast to this, the phylogenetic evidence for the complex V (A-ATPase) suggest that only four genes (*atpA*, *atpB*, *atpC*, *atpD*) were transfer from archaea-to-bacteria, while the remaining genes (*atpE*, *atpF*, *atpI*, *atpK*) do not possess bacterial homologs.

Interestingly, the observed monophyly of all Halobacteriales proteins in the trees might suggest their acquisition at the ancestor of this halophilic group, what has been shown in a previous study (). However, the clustering of genes that belong to the operon-like structures detected in this study shows that five genes (*nuoJ*, *ISP*, *coxB*, *atpK*, and *atpE*), within four complexes (I, III, IV, and V), are split in ≥2 protein families (see Table 8.1). From these cases, only four (*nuoJ, coxB, atpK*, and

Table 8.1. Genes involved in the aerobic respiration chain system in Halobacteriales genomes and their categories based on presence or absence of HGTs.

Complex	*Complex*	*Gene*	*Total* (50)	*S*	*I*	*E*	*U*	*Gene family*
Complex I	NADH	*nuoN*	1				1	1268
	dehydrogenase	*nuoM*	1				1	760
		nuoL	1				1	681
		nuoK	1				1	1430
		nuoJ	4	4				23825 | 7346 | 10131 | 2409
		nuoI	1				1	1564
		nuoH	1				1	407
		nuoCD	1				1	924
		nuoB	1				1	416
		nuoA	1		1			1500
Complex II	Succinate	*sdhD*	1	1				2462
	dehydrogenase	*sdhC*	1		1			2448
		sdhB	1				1	974
		sdhA	1				1	218
Complex III		*ISP*	7	2	3		2	10273 | 9130 | 2788 | 19923 | 5015 | 5120 | 3166
		CytC	1	1				3180
		CytB	1		1			3181
Complex IV	Cytochrome c	*coxD*	1	1				6126
	oxidase	*coxC*	1				1	2774
		coxB	2		2			2969 | 3462
		coxA	1				1	1940
		coxAC	1		1			3017
	Cytochrome bc	*cydB*	1				1	4574
	oxidase	*cydA*	1				1	2832
Complex V	A-type ATPase	*atpK*	4	4				25375 | 17534 | 14006 | 7389
		atpI	1			1		1015
		atpF	1	1				1075
		atpE	6	6				25553 | 22790 | 25376 | 17929 | 11965 | 1063
		atpD	1			1		38
		atpC	1			1		1055
		atpB	1			1		33
		atpA	1			1		34

Gene family category: *S, Specific; I, Import; E, Export; U, Undefined transfer*

atpE), have the same gene classification in the context of HGT (import or archaeal-specific), whereas the gene *ISP* has ≥2 annotations (archaeal-specific, import or undefined transfer). Similar cases occurred the previous analysis of the Methanosarcinales group (Chapter 7).

8.4 Comparison of the findings of this study and those of Nelson-Shathi et al. (2012).

In a previous study, Nelson-Sathi et al. (2012) reported a set of 213 genes involved in four out of the five the aerobic respiratory chain complexes (I, II, II, and IV) across 10 Halobacteriales genomes (named hereafter as "set A"). In the present study, 1,179 genes representing the five complexes of the aerobic respiratory system were identified across 23 Halobacteriales genomes (named hereafter as "set B"), and from this set, 675 genes were identified as part of an operon-like structure. In order to assess whether or not the genes reported by Nelson-Sathi are part of the like-operon structures identified in this study, both sets of genes (A and B) were compared. Importantly, 31 genes from Nelson-Sathi's work were excluded because they were either not found across most 23 genomes (*FrdA*, *COX15*, *NdhF*) or were not identified in the current NCBI database (*CoyE*, *CoxA*, *NuoF*, *CydA*, and *CydSB*).

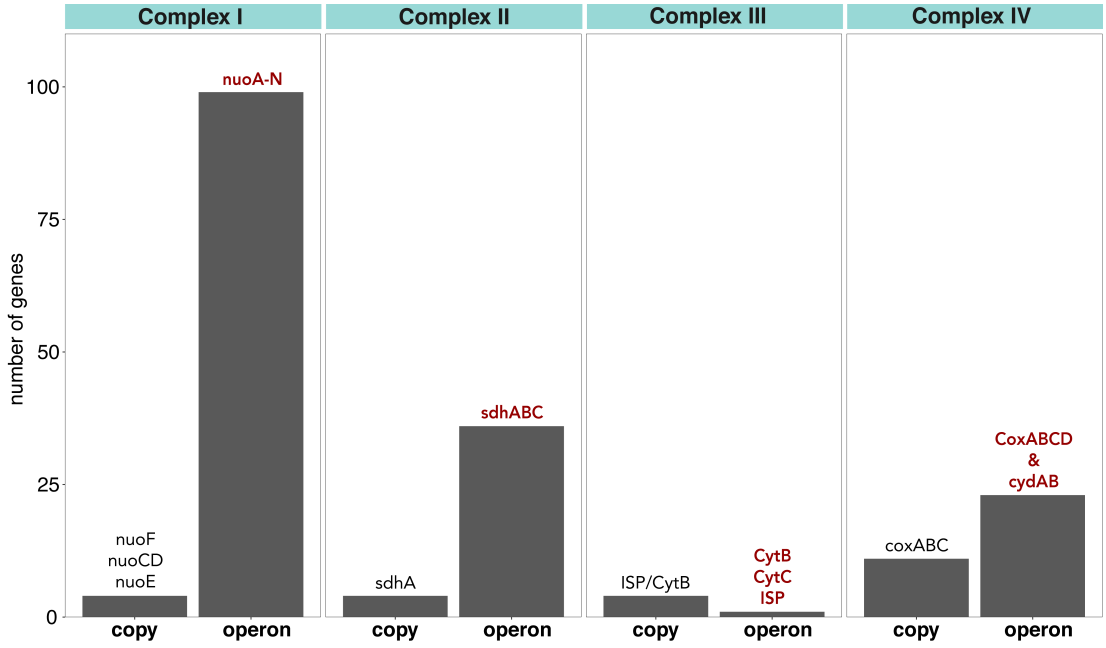


Figure 8.6. Classification of genes as belonging to operon-like structures or as gene copies involved in the aerobic respiratory chain system from Nelson-Sathi et al. (2012). Genes identified as part of the operon-like structures (legends in black at their corresponding bar plot), and cases defining gene copies (legends in red at their corresponding bar plot) for each membrane complex were identified (top headers in grey color). The number of genes composing each complex is as follow: *complex I*, 11 genes; *complex II*, 4 genes; *complex III*, 3 genes; *complex IV*, 5 genes (*coxABC* and *cydAB*). Genes reported by Nelson-Sathi et al. (2012) were obtained in the supplementary material from the publication.

The comparative analysis of the remaining 182 genes from the Nelson-Sathi's study shows that only 159 genes are located in the predicted operons, while 23 genes were identified as copies in other regions of the genome. Most of these gene copies correspond to the complex IV (11 copies), while the other complexes (I, II and III) posses 4 copies each. Accordingly, complex II, III and IV are found be the ones holding a considerable number of gene copies (e.g., iron sulfur protein (ISP) from complex III, as well as *coxA*, *coxB*, and *coxC* from complex IV), which can belong to the same or to a different protein family. From these cases HGTs were detected and, specficially, they were classified as imports (11), exports (2) and undefined transfers (10). Amog these cases, *coxC* was designed as export, whereas *nuoE*, *nuoF*, *sdhA*, *ISP*, *coxA* and *coxC*. However, after a manual curation of the gene trees, these results were in agreement with those from Nelson-Sathi's study.

8.5 Discussion of the chapter: The respiratory complexes in Halobacteriales were acquired from bacteria and their operon-like organization has been well conserved.

Earlier studies unveiled the presence of HGT events in the evolution of Halobacteriales (Ng et al. 2000; Boucher et al., 2003). More recently, some of the genes representing the aerobic respiration system in this archaeal clade has been suggested to be the outcome of HGT events from bacteria (Boucher et al., 2003; Nelson-Sathi et al. 2012). These complexes are usually organized in operons in Bacteria; however, such the operon organization has not been described in Halobacteriales yet. Here, I report the identification of five complexes involved with the respiratory chain system across 23 Halobacteriales genomes, in which represent 21 species. Furthermore, my findings show that a large

set of these genes clearly forms clusters in specific regions of the genome, which suggest an operon-like organization for all five complexes in the 23 genomes. The remaining genes correspond to paralogous counterparts from the five complexes, whose distribution was found largely widespread along the Halobacteriales genomes. Importantly, my findings confirm the acquisition of bacterial genes via HGT in four out of the five complexes (I, II, II and IV), whereas half of genes from the complex V (A-type ATPase synthase) appear to had be exported to bacteria and the other half of genes appear to be of archaean origin (i.e, archaeal specific).

First, my results show that the core number of subunits composing the complex I in Halobacteriales is 11 (nuoABCDHIJKLMN), which resembles its bacterial counterpart in cyanobacteria with the same number of subunits. However, the origin of complex I from Proteobacteria–where is composed by 14 subunits in one single operon structure–is supported by the phylogenetic evidence presented here. Thus, it is interesting to understand what has happened to the three missing genes in Halobacteria (*nuoE*, *nuoF*, and *nuoG*).

is supported by the phylogenetic evidence presented here. However, complex I in Proteobacteria is composed by 14 subunits in one single operon structure. Thus, it is interesting to understand what has happened to three missing genes in Halobacteria (*nuoE*, *nuoF*, and *nuoG*). The presence of these genes, known as the bacterial NADH dehydrogenase module or "module nuoEGF", were found found distantly located to the operon-like structure of complex I, i.e. not integrated in the operon, across all Halobacteriales.

The incomplete distribution of the acquired bacterial module nuoEGF in Halobacteriales is consistent with a previous report (Friedrich and Scheide 2000). Likewise, my results show that two subunits of the complex III (*cytC* and *cytB*) were absent in 4 Halobacteriales. Also, subunits of the 'cytochrome C oxidase' (*coxA*, *coxC*, *coxD*) or 'cytochrome bd oxidase' (*cydA*, *cydB*) of the complex IV were missed mainly in members of the genera *Halobacerium*, *Halakalicoccus*, *Natrialba*, *Haloterrigena*, *Halopiger*, *Natromonas*, *Haloarcula* and *Haloquadratum*, while both oxidases are retained in other species. It is well known that the composition of some of the five respiratory complexes can vary among domains or groups (Friedrich and Scheide, 2000), thus, it is also clear that they have been subject of major genomic rearrangements through the evolution of Halobacteriales (Friedrich and Scheide 2000).

Second, my results have shown that there are multiple gene copies of the membrane complexes located widespread along the Halobacteriales genomes, although some of these paralogs, mainly from complex IV, can form additional operons (complete and incomplete). According to this, a previous report has shown that operons are *created* and *destroyed* continuously, so that several paralogous complete or incomplete operons can be detected in the genome (Price et al. 2006). Price and coauthors even suggest that some of these "incomplete" clusters of genes represent the final state of the operon (Price et al. 2006). In any case, if these copies are functional, it is still unclear what could be their main role main role in the respiratory chain. Also, it is important to recognize that the considerable proportion of these paralogous genes might affect the identification of the complete respiratory complexes, as we all it genomic organizational structure. For example, the number of putative orthologs through the reciprocal best BLAST hits (rbbhs) approach was set to one rbbhs to avoid two methodological issues: **1)** a *high number paralogous genes that might disturb the phylogenetic signal* (Struck 2013), and **2)** protein families with an excessive number of paralogous proteins could increase significantly the computational time in the reconstruction of each phylogenetic tree. The identification of paralogous operons in a genome can represent an important challenge when the evolutionary

history of a specific operon wants to be reconstructed. Usually, the presence of paralogs is associated or confounded only with differential gene loss; however, it might be also influenced by HGT events (Boucher et al., 2003), and other paralogs or duplications among close related species (Bratlie et al. 2010; Correia et al. 2003; Xie et al. 2003).

Another important problem introduced by the presence of a high number of paralogous proteins is the clustering of such copies in different gene families, which might lead to different evolutionary histories despite of having the same function. For instance, some copies of a particular respiratory complex are annotated as archaeal specific (e.g. *nuoJ*, *atpH*), while other copies are classified differently according to our definitions: as "archaeal-specific" and "exports" for *atpI*, or as "archaeal-specific" and "undefined transfer" for *nuoM*, *atpD*, or as "imports" and "undefined transfers" for *nuoK*, *nuoL*, *coxA*, *coxB*, or as "export" and "undefined transfer" for coxC. The conflicting classification of some of these does not contradict the main results of this thesis and of previously reported findings (Nelson-Sathi et al. 2012; Nelson-Sathi et al. 2015). Rather, this paralogy-annotation problem might affect the classification of the direction of the HGT events (i.e., undefined transfer); for instance, on whether an archaeal specific gene was probably transferred later to bacteria (archaeal-specific and export), or in misguiding the interpretation of their evolution as imports. Thus, a manual inspection of each gene tree was essential to distinguish and correct these complex cases. Previous reports have already adviced on the potential problems caused by gene copies in the reconstruction of gene families and phylogenies (Altenhoff et al. 2012; Koonin and Galperin 2003; Roy 2009; Struck 2013). Therefore, future large-scale phylogenomic analysis must take into account the presence of gene copies, particularly in those genomes, such as Halobacteriales, that contain highly divergent genes with the same function (Boucher et al. 2004; Papke et al. 2007).

It has been largely suggested that the aerobic respiratory chain system in Archaea was acquired from Bacteria. Findings over 10 Halobactaeriales genomes supporting such idea were recently reported by Nelson-Sathi et al. (2012), and they further suggest that the acquisition of hundreds of genes from bacteria to the ancestor of Halobacteriales, including this aerobic respiratory chain system, occurred in a single HGT event. Despite the high number of gene copies and the presence of some paralogous operons, however, my findings demonstrate that all genes belonging to the aerobic respiration chain complexes I, II, II, and IV were acquired from bacteria in a single event at the ancestor of Halobacteriales. It is important to underline that this analysis was based on the context of 23 Halobacteriales genomes, which represent 21 different species from diverse environments and, importantly, they are not over-represented by haloarchaea strains. Based on the analysis performed in this study, the following genome mechanisms have been clearly involved in the evolution of the respiratory membrane-bound complexes, and should be also considered in future analysis to avoid methodological problems: 1) the insertion of foreign genes via HGT, 2) gene duplication, and 3) genome rearrangements. In the particular case of HGT, the insertion of particular genes or complete operons could be inserted in any part of the genome have been previously exposed in other systems (Bratlie et al. 2010; Price et al. 2006; Xie et al. 2003).

9

CONCLUSIONS

During their evolution, prokaryotes have been profoundly exposed to multiple and pervasive HGT events, creating highly connected networks among numerous groups of close and distant related species. In Archaea, bacterial acquisitions have provided genomic diversity, gain of some features to adapt and explore new environments, and play a central role in the formation of prokaryotic groups (Boto 2010; Koonin et al. 2001; Swithers et al. 2012; Wiedenbeck and Cohan 2011). In this study, I investigated the evolution of Archaea in the context of bacterial gene acquisitions via HGT using a large dataset of 1,981 prokaryotic genomes. The aim of this research was focused on study the influence of HGT events Archaea and Bacteria on three different subjects of interest: 1) origin of higher taxonomic archaeal clades, 2) evolution of the energy and methane metabolisms in Methanosarcinales and its influence on the transitions between freshwater and saltwater environments, and 3) the origin and evolution in operons of the aerobic respiratory system in Halobacteriales.

The phylogenetic reconstruction of HGT events between archaea and bacteria, followed by their quantification and characterization in this study, shows that interdomain gene transfers are highly asymmetric, so that transfers from bacteria-to-archaea (*imports*) are more than fivefold more frequent than those from archaea-to-bacteria (*exports*). In contrast, there is a large fraction of genes that represent an archaeal origin because bacterial homologs were not detected (the so called *archaeal-specific*). From those bacterial acquisitions, a considerable amount of imports are detected exclusively in specific archaeal clades (imports clade specific). The comparative analysis of the tree branches from both set of genes A and B suggests that “archaeal-specific genes” (set A) and “imports” (set B) evolved along the same phylogeny, therefore, these bacterial acquisitions can be traced back to the origin of the 13 traditionally recognized archaeal taxa. In addition, the evidence shows that multiple metabolic functions from bacteria were implicated as key metabolic innovations during the major evolutionary transitions via HGT (Nelson-Sathi et al. 2015).

By taking in account 10 complete genomes from **Methanosarcinales**, my research confirmed that aournd 10%–35% of their genomes have been subject to HGT. The highest percetange of HGTs corresponds to species of the Methanosarcinaceae family, and while the lowest fraction of HGTs is found in the Methanosaetaceae family. Furthermore, my results confirmed that particular bacterial groups, such as Firmicutes, Proteobacteria and Clostridia, are the main gene donors of the HGTs detected, given that they show the highest frequency among other members of bacteria. These findings are in agreement with previous studies (Allen et al. 2009; Deppenmeier et al. 2002; Fournier and Gogarten 2008; Garushyants et al. 2015). My major interest in this study was to know whether the metabolic versatility observed in the methane and energy metabolisms of Methanosarcinales was influenced by HGT events from bacteria. Two main findings were obtained from my analyses. First, several bacterial acquisitions were found be involved in the metabolic expansions that originated the aceticlastic pathway. The hypothesis on the bacterial origin of the methylotrophic pathway remains

unclear. On the one hand, 11 out of 12 methyltransferases (MTs) involved in this pathway are very likely bacterial acquisition; however, this assumption is difficult to support because the transfers in these phylogenies are not readily polarized since the distribution of MTs is restricted to Methanosarcinales and Methanobacteriales in Archaea, and highly restricted to different species in few groups of Bacteria. Although a recent analysis on the distribution of MTs along Archaea suggests their presence in other new clades (Vanwonterghem et al. 2016), such as Methanofastidiosales, Verstraetearchaeota or Bathyarchaeota, still highly restricted.

The second finding point out to the observation that multiple species in Methanosarcinales have loss the capability of growing under hydrogenotrophic conditions, in particular in the so-called *versatile methanogens* of the Methanosarcina genus. My findings unveil that major genes losses have occurred in the electron transport chain and energy conservation; probably, as probable replacement of the two key protein complexes (hydrogenases) of the hydrogenotrophic pathway in some Methanosarcinales genomes. For instance, evidence suggests that the hydrogenase Ech was replaced by Rnf complex acquired from bacteria via HGT followed by a secondary bacterial acquisition (the membrane complex Mrp), which make possible, as consequence, the transitions between freshwater and marine environments in Methanosarcinales. Both bacterial acquisitions and this aquatic transition might be responsible, in part, of dividing methanogens in Methanosarcinales as electron source H_2-dependent and H_2-independent. This aquatic transition expanded the niche in Methanosarcinales from an environment with highly competitiveness for methanogenic substrates to perform methylotrophic methanogenesis (freshwater) to an environment where there is no such competitiveness (marine). Overall, HGT events, gene losses, and environmental changes have shaped together the evolution of the metabolic versatily in Methanosarcinales.

In agreement with Nelson-Sathi et al. (2012), my findings confirm that most complexes of the aerobic respiration chain system in **Halobacteriales** have been acquired from bacteria via HGT. This bacterial acquisition occurred in four out of five complexes (complexes I, II, III, and IV) encompassing this aerobic respiratory system. In contrast, the phylogenetic gene trees of the complex V suggest that some its genes were transferred to bacteria. Similarly as in Bacteria, this acquired respiration chain system involved a large number of proteins, that are conserved and organized in operon-like manner among the 23 Halobacteriales genomes studied here. However, the location of each individual operon along the genome is, apparently, randomly positioned in each of the members in Halobacteriales. In addition, these findings were compared with a list of candidate genes belonging to the aerobic respiratory chain in Halobacteriales previously published (Nelson-Sathi et al. 2012). The comparison revealed that most of the published genes (75%) correspond to genes that are conserved and organized in an operon-like manner. In the light of these results, future hypotheses on the evolution of the aerobic respiratory chain system in Halobacteriales should consider to evaluate the genomic organization in an operon-like manner of the five membrane protein complexes. This is because the genes forming the five membrane-bound complexes have several gene copies that do not form complete like operon structure or that might be part of other operon and metabolic pathway.

Overall, this research opens new insights on the strong influence that the interdomain lateral gene transfer has had in the origin and evolution of Archaea at different hierarchical levels. Particularly, we were able to recover a global well-resolved picture of the continuous acquisition and crucial evolutionary transitions in two of the most extraordinary and inventive groups of Archaea, the Methanosarcinales and Halobacteriales. These findings, undoubtedly, were achieved with enthusiasm and constant surprises along the entire research.

REFERENCES

Achtman M, Wagner A. 2008. Microbial diversity and the genetic nature of microbial species. *Nature Rev Microbiol* **6**: 431-440.

Akiba T, Koyama K, Ishiki Y et al. 1960. On the mechanism of the development of multiple-drug-resistant clones of Shigella. *Jpn J Microbiol* **4**: 219-227.

Allen MA, Lauro FM, Williams TJ et al. 2009. The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: the role of genome evolution in cold adaptation. *ISME J* **3**: 1012-1035.

Altekar W, Rangaswamy V. 1992. Degradation of endogenous fructose during catabolism of sucrose and mannitol on halophilic archaebacteria. *Arch Microbiol* **158**: 356-363.

Altenhoff AM, Studer RA, Robinson-Rechavi M, Dessimoz C. 2012. Resolving the ortholog conjecture: orthologs tend to be weakly, but significantly, more similar in function than paralogs. *PLoS Comput Biol* **8**(5): e1002514.

Altschul SF, Madden TL, Schaffer AA et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**(17): 3389-3402.

Amils R. 2011. Euryarchaeota. In *Encyclopledia of Astrobiology*, (ed. M Garfaud, R Amils, JC Quintanilla, HJ Cleaves II, WM Irvine, DL Pinti, M Viso), p. 515. Springer Berlin Heidelberg, Madrid, Spain.

Angelidaki I, Karakashev D, Batstone DJ et al. 2011. Biomethanation and its potential. In *Methods in Enzymology*, Vol 494. Methods in Methane metabolism, Part A (ed. AC Rosenzweig, SW Ragsdale). Elseiver Inc., London, United Kingdom.

Baker BJ, Comolli JR, Dick GJ et al. 2010. Enigmatic, ultrasmall, uncultivated Archaea. *Proc Natl Acad Sci USA* **107**(19): 8806-8811.

Bapteste E, Brochier C, Boucher Y. 2005. Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea* **1**: 353–363.

Barns SM, Delwiche CF, Palmer JD, N.R. P. 1996. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci USA* **93**(17): 9188-9193.

Barns SM, Fundyga RE, Jeffries MW, Pace NR. 1994. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc Natl Acad Sci USA* **91**: 1609-1613.

Becker EA, Seitzer PM, Tritt A et al. 2014. Phylogenetically driven sequencing of extremely halophilic Archaea reveals strategies for static and dynamic osmo-response. *PLoS Genet* **10**(11): e1004784.

Berger S, Welte C, Deppenmeier U. 2012. Acetate activation in Methanosaeta thermophila: Characterization of the key pyrophosphatase and acetyl-CoA synthetase. *Archaea*: 2012:315153.

Bianchi ME, Radding CM. 1983. Insertions, deletions and mismatches in heteroduplex DNA made by recA protein. *Cell* **35**: 511–520.

Boone DR, Mathrani IM, Liu Y, Menaia JAGF, Mah RA, Boone JE. 1993. Isolation and characterization of *Methanohalophilus portucalensis* sp. nov. and DNA reassociation study of the genus *Methanohalophilus*. *Int J Syst Bacteriol* 43:430–437.

Boone DR. 2001. Genus V. Methanolobus Konig and Stetter 1983, 439VP (Effective publication: König and Stetter 1982, 488). In: Boone DR, Castenholz RW, Garrity GM (eds) Bergey's manual of systematic bacteriology, The Archaea and the deeply branching and phototrophic Bacteria, vol 1, 2nd edn. Springer, New York, pp 283–287.

Boone DR, Baker CC. 2001. Genus VI. *Methanosalsum* gen. nov. In: Boone DR, Castenholz RW, Garrity GM (eds) *Bergey's manual of systematic bacteriology*, The Archaea and the deeply branching and phototrophic Bacteria, vol 1, 2nd edn. Springer, New York, pp 287–289.

Borrel G, Gaci N, Peyret P et al. 2014. Unique characteristics of the pyrrolysine system in the 7th order of methanogens: implication for the evolution of a genetic code expansion cassette. *Archaea* **374146**: 1-12.

Borrel G, O'Toole PW, Harris HSR et al. 2013. Phylogenomic data support a seventh order of methylotrophic methanogens and provide insights into the evolution of methanogenesis. *Genome Biol Evol* **5**(10): 1769–1780.

Boto L. 2010. Horizontal gene transfer in evolution: facts and challenges. *Proc R Soc B* **277**(1683): 277, 819–827.

Boucher Y, Douady CJ, Papke T et al. 2003. Lateral gene transfer and the origin of prokaryotic groups. *Ann Rev Genet* **37**: 283-328.

Boucher Y, Douady CJ, Sharma AK et al. 2004. Intragenomic heterogeneity and intergenomic recombination among Halorchaeal rRNA genes. *J Bacteriol* **186**(12): 3980-3990.

Boucher, Y., Douady, C.J., Papke, T., Walsh, D.A., Boudreau, M. E. R., Nesbø, C.L., Case, R.J., Doolittle, W. F. 2003. Lateral gene transfer and the origin of prokaryotic groups. *Ann. Rev. Genet.* 37:283-328.

Bock, AK, Schönheit P. 1995. Growth of *Methanosarcina barkeri* (Fusaro) under nonmethanogenic conditions by the fermentation of pyruvate to acetate: ATP synthesis via the mechanism of substrate level phosphorylation. *J. Bacteriol.* 177(8): 2002–2007.

Bratlie MS, Johansen J, Drabløs F. 2010. Relationship between operon preference and functional properties of persistent genes in bacterial genomes. *BMC Genomics* **11**(71): 22.

Brochier C, Philippe H, Moreira D. 2000. The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome. *Trends Genet* **16**: 529-533.

Brown JR. 2003. Ancient horizontal gene transfer. *Nat Rev Gen* **4**: 121-132.

Brügger K, Redder P, She Q et al. 2002. Mobile elements in archaea genomes. *FEMS Microbiol Lett* **206**(2): 131-141.

Buan NR, Metcalf WW. 2010. Methanogenesis by Methanosarcina acetivorans involves two structurally and functionally distinct classes of heterodisulfide reductase. *Mol Microbiol* **75**(4): 843-853.

Bueno E, Mesa S, Bedmar EJ et al. 2012. Bacterial adaptation of respiration from oxic to microoxic and anoxic conditions: redox control. *Antioxidants and Redox Signaling* **16**(8): 819-852.

Burggraff S, Stetter KO, Rouviere P. 1991. Methanopyrus kandleri: an archaeal methanoge unrelated to all other know methanogens. *Syst Appl Microbiol* **14**: 346–351.

Caro-Quintero A, Konstatinidis KT. 2015. Inter-phylum HGT has shaped the metaoblism of many mesophilic and anaerobic bacteria. *ISME J* **9**: 958-967.

Castelle CJ, Wrighton KC, Thomas BC et al. 2015. Genomic expansion of domain Archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr Biol* **25**: 690-701.

Cavalli-Sforza LL. 1998. The DNA revolution in population genetics. *Trends in Genet* **14**(2): 60-65.

Chaban B, Ng SYM, Jarrell KF. 2005. Archaeal habitats - from the extreme to the ordinary. *Can J Microbiol* **52**: 73–116.

Chistoserdova L, A. VJ, Lidstrom ME. 2005. A genomic view of methane oxidation by aerobic bacteria and anaerobic archaea. *Genome Biol* **6**: 208.

Chistoserdova L, Vorholt JA, Thauer RK, Lidstrom ME. 1998. C1 transfer enzymes and coenzymes linking methytrophic bacteria and methanogenomic archaea. *Science* **281**: 99-102.

Chivian D, Brodie EL, Alm EJ et al. 2008. Environmental genomics reveals a single-species ecosystem deep within earth. *Science* **322**: 275-278.

Ciccarelli FD, Doerks T, von Mering C et al. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**(5765): 1283-1287.

Clingenpeel S, Kan J, Macur RE et al. 2013. Yellowstone Lake Nanoarchaeota. *Fromtiers in Microbiology* **4**: 274.

Cohan FM. 2011. Are species cohesive? – A view from bacteriology. In *Bacterial population genetics: A tribute to Thomas S Whittam*, (ed. S Walks, P Feng), pp. 43 – 65. American Society for Mocrobiology Press, Washington, DC, USA.

Conrad R. 1999. Contribution of hydrogen to methane production and control of hydrogen concentrations in methanogenic soils and sediments. *FEMS Microbiol* **28**: 193-202.

Cordero OX, Hogeweg P. 2009. The impact of long-distance horizontal gene transfer on prokaryotic genome size. *Proc Natl Acad Sci USA* **106**(51): 21748 –21753.

Correia FF, Plummer AR, Ellen RP et al. 2003. Two paralogous families of a two-gene subtilisin operon are widely distributed in oral treponemes. *J Bacteriol* **185**(23): 6860-6869.

Cox CJ, Foster PG, Hirt RP et al. 2008. The archaebacterial origin of eukaryotes. *Proc Natl Acad Sci USA* **105**: 20356–20361.

Dagan T. 2011. Phylogenomics networks. *Trends Microbiol* **19**(10): 483-491.

Daubin, V., Gouy, M., Perrière, G. 2002. A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res.* 7(12): 1080-1090.

Dagan T, Artzy-Randrup Y, Martin W. 2008. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc Natl Acad Sci USA* **105**(29): 10039-10044.

Dagan T, Martin W. 2006. The tree of one percent. *Genome Biol* **7**: 118.

Darmon E, Leach DRF. 2014. Gacterial Genome Instability. *Microbiol Mol Biol Rev* **78**(1): 1-39.

de Poorter LM, Geerts WJ, Keltjens JT. 2007. Coupling of *Methanothermobacter thermautotrophicus* methane formation and growth in fed-batch and continuous cultures under different H2 gassing regimens. *Appl Environ Microbiol* **73**: 740-749.

Deppenmeier U. 1995. Different structure and expression of the operons econding the membrane-bound hydrogenases from *Methanosarcina mazei* Gö1. *Arch Microbiol* **164**: 370-376.

Deppenmeier U. 2002. The unique biochemistry of methanogenesis. *Prog Nucleic Acid Res Mol Biol* **71**: 223-283.
Deppenmeier U, Johann A, Hartsch T et al. 2002. The genome of *Methanosarcina Mazei*: Evidence for Lateral Gene Transfer between Bacteria and Archaea *J Mol Microbiol Biotechnol* **4**(4): 453-461.
Dibrova DV, Galperin MY, Mulkidjanian AY. 2010. Characterization of the N-ATPase, a distict, laterally transferred Na+-translocating form of the bacterial F-type membrane ATPase. *Bioinformatics (Oxford, England)* **26**(12): 1473-1476.
Doerfert SN, Reichlen M, Iyer P, Wang M, Ferry JG. 2009. *Methanolobus zinderi* sp. nov., a methylotrophic methanogen isolated from a deep subsurface coal seam. *Int J Syst Evol Microbiol* 59:1064–1069.
Doolittle F. 2000. Uprooting the Tree of Life. *Scientific American* **282**: 90-95.
Doolittle F, Papke T. 2006. Genomics and the bacterial species problem. *Genome Biol* **7**: 116.
Dubey GP, Ben-Yehuda S. 2011. Intercellular nanotubes mediate bacterial communication. *Cell* **144**: 590-600.
Elberson MA, Sowers KR, 1997. Isolation of an aceticlastic strain of *Methanosarcina siciliae* from marine canyon sediments and emendation of the species description for *Methanosarcina siciliae*. *Int J Syst Evol Microbiol* 47: 1258-1261.
Ehrhardt CJ, Haymon RM, Lamontagne MG, Holden PA. 2007. Evidence for hydrothermal Archaea within the basaltic flanks of the East Pacific Rise. *Environ Microbiol* **9**: 900–912.
Elkins JG, Podar M, Graham DE et al. 2008. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc Natl Acad Sci USA* **105**: 8102–8107.
Embley TM, Williams TA. 2015. Evolution: Steps on the road to eukaryotes. *Nature* **521**: 169–170.
Enpadinhas N, da Costa MS. 2008. Osmoadaptation mechanisms in prokaryotes: distribution of compatible solutes. *Int Microbiol* **11**: 151-161.
Evans PD, Mekel-Bobrov N, Vallender EJ et al. 2006. Evidence that the adaptive allele of the brain size gene microcephalin introgressed into *Homo sapiens* from an archaic *Homo* lineage. *Proc Natl Acad Sci USA* **103**: 18178-18183.
Evans PN, D.H. P, Chadwicks GL et al. 2015. Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**(6259): 434-438.
Falb M, Müller K, Königsmaier L et al. 2008. Meabolism of halophilic archaea. *Extremophiles* **12**: 177-196.
Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool* **27**: 401–410.
Fendrihan S, Legat A, Pfaffenhuemer M et al. 2006. Extremely halophilic archaea and the issue of long-term microbial survival. *Rev Environ Sci Biotechnol* **5**(2-3): 203-218.
Ferry JG. 1999. Enzymology of one-carbon metabolism in methanogenic pathways. *FEMS Microbiol Rev* **23**(1): 13-38.
Ferry JG. 2002. Methanogenesis biochemstry. *eLS*: 1-9.
Ferry JG. 2011. Fundamentals of methanogenic pathways that are key to the biomethanation of complex biomass. *Curr Opin Biothecn* **22**: 351-357.
Ferry JG, Lessner DJ. 2008. Methanogenesis in marine sediments. *Ann N Y Acad Sci* **1125**: 147-157.
Filée J, Siguier P, Chandler M. 2007. Insertion sequence diversity in Archaea. *Microbiol Mol Biol Rev* **71**(1): 121-157.
Finn RD, Bateman A, Clements J et al. 2014. The Pfam protein families database. *Nucleic Acids Res* **42**(D1): D222-D230.
Fournier GP, Dick AA, Williams D, Gogarten JP. 2011. Evolution of the archaea: emerging views on origins and phylogeny. *Res Microbiol* **162**(1): 92-98.
Fournier GP, Gogarten JP. 2008. Evolution of acetoclastic methanogenesis in Methanosarcina via Horizontal Gene Tranfer from cellulolytic Clostridia *J Bacteriol* **190**(3): 1124–1127.
Franco AA. 2004. The Bacteroidetes fragilis pathogenicity island is contained in a putative novel conjugative transposons *J Bacteriol* **186**: 6077-6092.
Franzmann PD, Springer N, Ludwig W, Conway De Macario E, Rohde M. 1992. A methanogenic archaeon from Ace Lake, Antarctica: *Methanococcoides burtonii* sp. nov. *Syst Appl Microbiol* 15:573–581.
Fraser C, Alm EJ, Polz MF et al. 2008. The bacterial species challenge: making sense of genetic and ecological diversity. *Science* **323**: 741-746.
Friedrich T, Scheide D. 2000. The respiration complex I of bacteria, archaea and eukarya and its module common with membrane-bound multisubunit hydrogenases. *FEBS Lett* **479**: 1-5.
Gal-Mor O, Finlay B. 2006. Pathogenicity islands: a molecular toolbox for bacterial virulence. *Cell Microbiol* **8**(11): 1707-1719.
Galagan JE, Nusbaum C, Roy A et al. 2002. The genome of M. acetivorans reveals extensive metabolic and physiological diversity. *Genome Res* **12**: 532-542.
Garcia JL, Patel BK, Ollivier B. 2000. Taxonomic, phylogenetic, and ecological diversity of methanogenic Archaea. *Anaerobe* **6**: 205-226.

García-Vallve S, Romeu A, Palau J. 2000. Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res* **10**: 1719-1725.
Garcia-Vallvé S, Simó FX, Montero MA et al. 2002. Simultaneous horizontal gene transfer of a gene coding for ribosomal protein L27 and operational genes in Arthrobacter Sp. *J Mol Evol* **55**: 632-637.
Garrett RA, Klenk HP. 2007. *Archaea: Evolution, Physiology, and Molecular Biology*. Blackwell Publishing Ltd.
Garrity GM, Holt JG. 2001a. Phylum AI. Crenarchaeota phy. nov. p.169. In *Bergey's Manual of Systematic Bacteriology*, Vol 1 (ed. DR Boone, R Castenholz, GM Garrity). Springer, New York, US.
Garrity GM, Holt JG. 2001b. Phylum AII. Euryarchaeota phy. nov., pp.169. In *Bergey's Manual of Systematic Bacteriology*, Vol 1 (ed. DR Boone, R Castenholz, GM Garrity). Springer, New York, US.
Garushyants S, Kazanov MD, Gelfand MS. 2015. Horizontal gene transfer and evolution in Methanosarcina. *BMC Evol Biol* **15**: 102.
Goldenfeld N, Woese C. 2007. Biology's next revolution. *Nature* **445**: 369.
Gough J, Karplus K, Hughey R, Chothia C. 2001. Assignment of homology to genome sequences using a library of Hidden Markov Models that represent all proteins of known structure. *J Mol Evol* **313**(4): 903-919.
Greening C, Ahmed FH, Mohamed E et al. 2016. Physiology, Biochemistry, and application of F420- and Fo-dependent redox reactions. *Microbiol Mol Biol Rev* **80**(2): 451-493.
Gressmann H, Bodo L, Ghai R et al. 2005. Gain and loss of multiple genes during their evolution of *Helicobacter pyroli*. *PLoS Genet* **1**(4): e43.
Gribaldo S, Brochier-Armanet C. 2006. The origin and evolution of Archaea: a state of the art. *Philos Trans R Soc Lond B Biol Sci* **361**: 1007-1022.
Griffiths JFA, Miller JH, Suzuki DT et al. 2000. *An Introduction to Genetic Analysis*. W. H. Freeman, New York, USA.
Grigoriev I, Nordberg H, Shabalov I et al. 2012. The Genome Portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res* **40**(Databases issue): D26-32.
Guest RL, Wang J, Wong JL et al. 2017. A bacterial stress response regulates respiratory protein complexes to control envelope stress adaptation. *J Bacteriol* **199**(20): e00153-00117.
Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696-704.
Guss AM, Kulkarni G, Metcalf WW. 2009. Differences in hydrogenase gene expression between *Methanosarcina acetivorans* and *Methanosarcina barkeri*. *J Bacteriol* **191**(8): 2826-2833.
L'Haridon L, Chalopin M, Colombo D, Toffin L. 1983. *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: 1978-1983.
Hartmann R, Sickinger HD, Oesterhelt D. 1980. Anaerobic growth of halobacteria. *Proc Natl Acad Sci USA* **77**: 3821-3825.
Hedderich R, Whitman WB. 2006. Physiology and biochemistry of the methane-producing archaea. In *Prokaryotes: a handboook on the biology of Bacteria*, (ed. S Falkow, E Rosenberg, KH Schleifer, E Stackebrandt, M Dworkin), pp. 1050-1079. Springer, New York, USA.
Hedderich R, Whitman WB. 2013. Physiology and biochemstry of the Methane-producing Archaea. In *The Prokaryotes Prokaryotic Physiology and Biochemistry*, (ed. E Rosenberg), pp. 635-662. Springer-Verlag Berlin Heidelberg, New Delhi, India.
Hedlund BP, Murugapiran SK, Alba TW et al. 2015. Uncultivated thermophiles: current status and spotlight on 'Aigarchaeota'. *Curr Opin Microbiol* **25**: 136–145.
Heiden S, Hedderich R, Setzke E, Thauer RK. 1994. Purification of a two-subunits cytochrome-b-containing heterodisulfide reductase from methanol-grow Methanosarcina barkeri. *Eur J Biochem* **221**: 855-861.
Hershberger KL, Barns SM, Reysenbach AL et al. 1996. Wide diversity of Crenarchaeota. *Nature* **380**: 420.
Hilario E, Gogarten JP. 1993. Horizontal transfer of ATPase genes – the tree of life becomes a net of life. *BioSystems* **31**: 111-119.
Hovey R, Lentes S, Ehrenreich A et al. 2005. DNA microarray analysis of Methanosarcina mazei Gö1 reveals adaptation to different methanogenic substrates. *Mol Gen Genomics* **273**: 225-239.
Huang J, Gogarten JP. 2006. Ancient horizontal gene transfer can benefit phylogenetic reconstruction. *Trends in Genet* **22**(7): 362-366.
Huber H, Hohn MJ, Rachel R et al. 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**: 63-67.
Huber H, Kreuter L. 2014. The phylum Nanoarchaeota. In *The Prokaryotes – Other Major Lineages of Bacteria and the Archaea*, (ed. E Rosenberg, EF DeLong, S Lory, E Stackebrandt, F Thompson). Springer-Verlag Berlin Heidelberg, Berlin, Germany.

Hunt BG, Ometto L, Wiurm Y et al. 2011. Relaxed selection is a precursor to the evolution of phenotypic plasticity. *Proc Natl Acad Sci USA* **108**(38): 15936-15941.

Hutten TJ, Bongaerts HCM, van der Drift C, Vogels GD. 1980. Acetate, methanol and carbon dioxide as substrates for growth of *Methanosarcina barkeri*. *Antonie van Leeuwenhoek J Microbiol Serol* **46**: 601-610.

Ito M, Morino M, Krulwich TA. 2017. Mrp antiporters have important role in diverse Bacteria and Archaea. *Front Microbiol* **8**: 2325.

Jain R, Rivera MC, Lake JA. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci USA* **96**: 3801–3806.

Jain R, Rivera MC, Moore JE, Lake JA. 2003. Horizontal gene transfer accelerates genome innovation and evolution. *Mol Biol Evol* **20**(10): 1598-1602.

Jasso-Chávez R, Apolinario EE, Sowers KR, Ferry JG. 2013. MrpA fucntions in energy conservation during the acetate-dependent growth of Methanosarcina acetivorans. *J Bacteriol* **195**: 3987-3994.

Jasso-Chávez R, Diaz-Perez C, Rodríguez-Zavala JS, Ferry JG. 2016. Functional role of mrpA in the mrpABCDEFG Na/H antiporter complex from the archaeon Methanosarcina acetivorans. *J Bacteriol* **199**(2): e00662-00616.

Jetten MSM, Stams AJM, Zehnder AJB. 1992. Methanogenesis from acetate: a comparison of the acetate metabolism in Methanothrix soengenii and Methanosarcina spp. *FEMS Microbiol* **88**: 181-197.

Johnsborg O, Eldholm V, Havarstein LS. 2007. Natural genetic transformation: prevalence, mechanisms and function. *Res Microbiol* **158**: 767-778.

Johnsen U, Selig M, Xavier KB et al. 2001. Different glycolytic pathways for glucose and fructose in the halophilic archaeon Halococcus saccharolyticus. *J Bacteriol* **186**: 6198-6207.

Junier T, Zdobnov EM. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics (Oxford, England)* **26**(13): 1669-1670.

Kadam PC, Ranade DR, Mandelco L, Boone DR. 1994. Isolation and characterization of *Methanolobus bombayensis* sp. nov., a methylotrophic methanogen that requires high concentrations of divalent cations. *Int J Syst Bacteriol* 44:603–607.

Kalyuzhnaya MG, Chistoserdova L. 2005. Community-level analysis: genes encoding methanopterin-depedent enzymes. *Methods in Enzymology* **397**: 443-454.

Kalyuzhnaya MG, Korotkova N, Crowther G et al. 2005. Analysis of gene islands involved in methanopterin-linked C1 transfer reactions reveals new functions and provides evolutionary insights. *J Bacteriol* **187**: 4607-4616.

Kalyuzhnaya MG, Lidstrom ME, Chistoserdova L. 2004. Utility of environmental primers targeting ancient enzymes: methylotrophs dectection in Lake Washington. *Microb Ecol* **48**: 463-472.

Kamagata Y, Kawasaki H, Oyaizu H, Nakamura K, Mikami E, Endo G, Koga Y, Yamasato K. 1992. Characterization of three thermophilic strains of *Methanothrix* (''*Methanosaeta*'') *thermophila* sp. nov. and rejection of *Methanothrix* (''*Methanosaeta*'') *thermoacetophila*. *Int. J. Syst. Bacteriol.* 42: 463–468.

Kanehisa M, Goto S. 2000. KEGG: Kyoto Ecyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**: 27-30.

Karrasch M, Börner G, Enssle M, Thauer RK. 1989. Formylmethanofuran dehydrogenase from methanogenic bacteria, a molybdoenzyme. *FEBS Lett* **253**(1-2): 226-230.

Kates M. 1993. Membrane lipids of archaea In *The biochemistry of Archaea (Archaebacteria)*, (ed. M Kates, DJ Kushner, AT Matheson), p. 261. Elsevier, New York, USA.

Katoh K, Frith M. 2012. Adding Unaligned Sequences Into an Existing Alignment Using MAFFT and LAST *Bioinformatics (Oxford, England)* **28**(23): 3144-3146.

Katoh K, Kuma K, Miyata T. 2001. Genetic algorithm-based maximum-likelihood analysis for molecular evolution. *J Mol Evol* **53**: 477–484.

Kauri T, Wallace R, Kushner DJ. 1990. Nutrition of the halophilic archaebacterium, *Haloferax volcanii*. *Syst Appl Microbiol* **13**: 14-18.

Kennedy, S. P., Ng, W.V., Salzberg, S. L., Hood, L., DasSarma, S. 2001. Understanding the adaptation of Haloobacterium species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res.* 11(10): 1641-1650.

Khomyakova M, Bükmez Ö, Thomas LK et al. 2011. A Methylaspartate Cycle in Haloarchaea. *Science* **331**(6015): 334-337.

Kletzin A, Heimerl T, Flechsler J et al. 2015. Cytochrome c in Archaea: distribution, maturation, cell architecture, and the special case of *Ignicoccus hospitalis*. *Front Microbiol* **6**: 439.

Konig H, Stetter KO. 1982. Isolation and characterization of *Methanolobus tindarius*, sp. nov., a coccoid methanogen growing only on methanol and methylamines. *Zbl Bakt Hyg I Abt Orig C* 3:478–490.

Koonin EV. 2009. Evolution of genome architecture. *Int J Biochem Cell Biol* **41**: 298-306.

Koonin EV, Galperin MY. 2003. *Sequence - Evolution - Function. Computational approaches in comparative genomics*. Springer US, US.
Koonin EV, Makarova AS, Aravind L. 2001. Horizontal gene transfer in prokaryotes: Quantification and classification. *Annu Rev Microbiol* **55**: 709-742.
Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic acids research* **36**(1): 6688-6719.
Koski LB, Morton RA, Golding GB. 2001. Codon bias an base composition are poor indicators of horizontally tranferred genes. *Mol Biol Evol* **18**: 404-412.
Kozubal MA, Romine M, Jennings RdM et al. 2013. Geoarchaeota: a new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park *ISME J* **7**: 622-634.
Kulkarni G, Kridelbaugh DM, Guss AM, Metcalf WW. 2009. Hydrogen is preferred intermediate in the energy-conserving electro transport chain of Methanosarcina barkeri. *Proc Natl Acad Sci USA* **106**(37): 15915-15920.
Kurland CG, Canback B, Berg OG. 2003. Horizontal gene transfer: A critical review. *Proc Natl Acad Sci USA* **100**(17): 9658–9662.
Lai MC, Gunsalus RP. 1992. Glycine betaine and potassium ion are the major compatlbie solutes in the extremely halophilic methanogen *Methanohalophilus strain* Z7302. *J Bacteriol* **174**(22): 7474-7477.
Lai MC, Hong TY, Gunsalus RP. 2000. Glycine betaine transport in the obligate halophilic archaeon *Methanohalophilus portucalensis*. *J Bacteriol* **182**(7): 5020-2024.
Lai MC, Sowers KR, Robertson DE et al. 1991. Distribution of compatible solutes in the halophilic methanogenic archaebacteria. *J Bacteriol* **173**(17): 5352-5358.
Lake JA. 1991. The order of sequence alignment can bias the selection of tree topology. *Mol Biol Evol* **8**: 378–385.
Lake JA, Rivera MC. 2004. Deriving the genomic tree of life in the presence of horizontal gene transfer: conditioned reconstruction. *Mol Biol Evol* **21**: 681-690.
Lane CE, Archibald JM. 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. *Trends Ecol Evol* **23**: 268–275.
Lang AS, Zhaxybayeva O, Beatty JT. 2012. Gene transfer agents: phage-like elements of genetic exchange. *Nat Rev Micro* **10**: 472–482.
Lang K, Schuldes J, Klingl A et al. 2014. New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genomic analysis of "Candidatus Methanoplasma termitum". *Appl Environ Microbiol* **81**: 1338 –1352.
Lawrence JG. 1997. Selfish operons and speciation by gene transfer. *Trends in Microbiol* **5**(9): 355-359.
Lercher MJ, Pál C. 2008. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol Biol Evol* **25**(3): 559–567.
Li Q, Li L, Rejtar T et al. 2006. Electron transport in the pathway of acetate conversion to methane in the marine archeon Methanosarcina acetivorans. *J Bacteriol* **188**(2): 702-710.
Lionel G, Kultima JR, Anderson GE. 2010. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics (Oxford, England)* **16**(18): 2334-2335.
Liu Y, Boone DR, Choy C. 1990. *Methanohalophilus oregonense* sp. nov., a methylotrophic methanogen from an alkaline, saline aquifer. *Int J Syst Bacteriol* 40:111–116.
Liu Y, Whitman WB. 2008. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann N Y Acad Sci* **1125**: 171-189.
Lomans BP, Maas R, Luderer R, Op den Camp HJM, Pol A, van der Drift C, Vogels GD. 1999. Isolation and characterization of *Methanomethylovorans hollandica* gen. nov., sp. nov., isolated from freshwater sediment, a methylotrophic methanogen able to grow on dimethyl sulfide and methanethiol. *Appl Environ Microbiol* 65, 3641–3650.
Lorenz MG, Wackernagel W. 1994. Bacterial gene-transfer by natural genetic-transformation in the environment. *Microbiol Rev* **58**: 563–602.
Lyimo TJ, Pol A, Op den Camp HJM, Harhangi HR, Vogels GD. 2000. *Methanosarcina semesiae* sp. nov., a dimethylsulfide-utilizing methanogen from mangrove sediment. *Int J Syst Evol Microbiol* 50:171–178
Lynch E, Langille M, Darling A et al. 2012. Sequencing of Seven Haloarchaeal Genomes Reveals Patterns of Genomic Flux. *PLoS One* **7**(7): e41389.
Ma K, Liu X, Dong. 2006. Methanosaeta harundinacea sp. nov., novel acetate-scavenging methanogen isolated from a UASB reactor. *Int. J. Syst. and Evol. Microbiol*. 56: 127-131.
Maeder DL, Anderson I, Brettin TS et al. 2006. The *Methanosarcina barkeri* genome: comparative analysis with Methanosarcina acetivorans and Methanosarcina mazei reveals extensive rearrangements within methanosarcinal genomes. *J Bacteriol* **188**(22): 7922-7931.

Maestrojuán GM, Boone DR. 1991. Characterization of *Methanosarcina barkeri* MST and 227, *Methanosarcina mazei* S-6T, and *Methanosarcina vacuolata* Z-761T. *Int J Syst Bacteriol* **41**(2): 267-274.

Maestrojuan GM, Boone JE, Mah RA, Menaia JAGF, Sachs MS, Boone DR. 1992 Taxonomy and Halotolerance of Mesophilic Methanosarcina Strains, Assignment of Strains to Species, and Synonymy of *Methanosarcina mazei* and *Methanosarcina frisia*. *Int J of Syst Evol Microbiol* 42: 561-567.

Mah RA. 1980. Isolation and characterization of *Methanococcus mazei*. *Curr Microbiol* **3**: 321-326.

Mahapatra A, Patel A, Soares JA et al. 2006. Characterization of a *Methanosarcina acetivorans* mutant unable to translate UAG as pyrrolysine. *Mol Microbiol* **59**(1): 56-66.

Makarova AS, Yutin N, Bell SD, Koonin EV. 2010. Evolution of diverse cell division and vesicle formation systems in Archaea. *Nat Rev Micro* **8**: 731-741.

Makarova KS, Koonin EV. 2003. Comparative genomics of archaea: how much have we learned in six years, and what's next? *Genome Biol* **4**: 115.

Makarova KS, Ponomarev VA, Koonin EV. 2001. Two C or not two C: recurrent disruption of Zn-ribbons, gene duplication, lineage-specific gene loss, and horizontal gene transfer in evolution of bacterial ribosomal proteins. *Genome Biol* **2**(9): RESEARCH 0033.0031-RESEARCH 0033.0014.

Makarova KS, Wolf YI, Alkhnbashi OS et al. 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Micro* **13**: 15.

Marteinsson VT, K. KJ, Kristmannsdóttir H et al. 2001. Discovery and Description of Giant Submarine Smectite Cones on the Seafloor in Eyjafjordur, Northern Iceland, and a Novel Thermal Microbial Habitat. *Appl Environ Microbiol* **67**: 827–833.

Martin FW, Baross J, Kelley D, Russell MJ. 2008. Hydrothermal vents and the origin of life. *Nat Rev Micro* **6**: 805-814.

Mayer F, Müller V. 2013. Adaptations of anaerobi archaea to life under extreme energy limitations. *FEMS Microbiol* **38**: 449-472.

Mochimaru H, Tamaki H, Hanada S, Imachi H, Nakamura K, Sakata S, Kamagata Y. 2009. *Methanolobus profundi* sp. nov., a methylotrophic methanogen isolated from deep subsurface sediments in a natural gas field. *Int J Syst Evol Microbiol.* 59:714-8.

McGlynn SE. 2017. Energy metabolism during anaerobic methane oxidation in ANME archaea. *Microbes Environ* **32**(1): 5-13.

McInerney JO, O'Connell ML, Pisani D. 2014. The hybrid nature of eukaryota and a consilent view of life on Earth. *Nature Rev Microbiol* **12**: 449-455.

Medini D, Donati C, Tettelin H et al. 2005. The microbial pan-genome. *Curr Opin Genet Dev* **15**(6): 589–594.

Meng J, Xu J, Qin D et al. 2014. Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. *ISME J* **8**(3): 650-659.

Meuer J, Kuettner HC, Zhang JK et al. 2002. Genetic analysis of the archeon Methanosarcina barkeri Fusaro reveals a central role for Ech hydrogenase and ferredoxin in methanogenesis and carbon fixation. *Proc Natl Acad Sci USA* **99**(8): 5632-5637.

Mevarech M, Werczberger R. 1985. Gene transfer in Halobacterium volcanii. *J Bacteriol* **162**: 461–462.

Miller TL, Woling MJ. 1985. Methanosphaera stadtmanae gen. nov., sp. nov.: a species that forms methane by reducing methanol with hydrogen. *Arch Microbiol* **141**: 116-122.

Min H, Zinder SH. 1989. Kinetics of acetate utilization by two thermophilic acetotrophic methanogens: *Methanosarcina sp.* strain cals-1 and *Methanotrhix sp.* strain cals-1. *Appl Environ Microbiol* **55**(2): 488-491.

Mojica FJ, Diez-Villasenor C, García-Martinez J, Soria E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**: 174–182.

Mori K, Lino T, Suzuki KI et al. 2011. Aceticlastic and NaCl-requiring methanogen "Methanosaeta pelagica" sp. nov., islolated from marine tidal flat sediment. *Appl Environ Microbiol* **78**(9): 3416-3423.

Müller JA, DasSarma S. 2005. Genomic analysis of aerobic respiration in the archaeon Halobacterium sp. strain NRC-1: Dimethyl sulfoxide and trimethylamine N-Oxide as terminal electron acceptors. *J Bacteriol* **187**(5): 1659-1667.

Nakamura Y, Matsuda H, Gojobori T. 2004. Biased biological functions of horizontallly transferred genes in prokaryotic genomes. *Nature* **36**: 760-766.

Naor A, Gophna U. 2013. Cell fusion and hybrids in Archaea: prospects for genome shuffling and accelerated strain development for biotechnology. *Bioengineered* **4**: 126–129.

Narayanan N, Krishnakumar B, Anupama VN, Manilal VB. 2009. Methanosaeta sp., the major archaeal endosymbiont of Metopus es. *Res Microbiol* **160**: 600-607.

Nealson K. 2008. A Korarchaeote yields genome sequencing. *Proc Natl Acad Sci USA* **105**(26): 8805 - 8806.

Nei M. 2003. *Mutation-Driven Evolution*. Oxford University Press, Pensylvania state University, United Kingdom.

Nelson-Sathi S, Dagan T, Janssen A et al. 2012. Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc Natl Acad Sci USA* **109**(50): 20537–20542.

Nelson-Sathi S, Sousa FL, Roettger M et al. 2015. Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature* **517**: 77–80.

Ney B, Ahmed FH, Carere CR et al. 2017. The methanogenic redox cofactor F420 is widely synthesized by aerobic soil bacteria. *ISME J* **11**: 125-137.

Ng WV, Kennedy S, P., Mahairas GG et al. 2000. Genome sequence of *Halobacterium* species NRC-1. *Proc Natl Acad Sci USA* **97**(22): 12176-12181.

Ni S, Boone DR. 1991. Isolation and characterization of a dimethyl sulfidedegrading methanogen, *Methanolobus siciliae* HI350, from an oil well, characterization of *M. siciliae* T4/MT, and emendation of *M. siciliae*. *Int J Syst Bacteriol* 41:410–416.

Ni S, Woese CR, Aldrich HC, Boone DR. 1994. Transfer of *Methanolobus siciliae* to the genus Methanosarcina, naming it *Methanosarcina siciliae*, and emendation of the genus Methanosarcina. *Int J Syst Bacteriol* 44:357–359.

Nuñez PA, Romero H, Farber MD, Rocha EPC. 2013. Natural selection for operons depends on genome size. *Genome Biol Evol* **5**(11): 2242-2254.

Nunoura T, Takaki Y, Kakuta J et al. 2011. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic acids research* **39**(8): 3204-3223.

Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299-304.

Olendzenzki L, Gogarten JP. 2009. Gene tranfer: Who benefits? In *Horizontal Gene Transfer: Genomes in flux*, (ed. MB Gogarten, JP Gogarten, L Olendzenzki), pp. 3-12. Humana Press, New York, USA.

Olsen GJ, Woese C. 1996. Lesson from an Archaeal genome: what are we learning from Methanococcus jannaschii? *Trends Genet* **12**(377–379).

Omelchenko MV, Makarova AS, Wolf YI et al. 2003. Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ. *Genome Biol* **4**: R55.

Oremland RS, Boone DR. 1994. *Methanolobus taylorii* sp. nov., a new methylotrophic, estuarine methanogen. *Int J Syst Bacteriol* 44:573–575.

Oren A. 1999. Bioenergetic aspects of Halophilism. *Microbiol Mol Biol Rev* **63**(2): 334-348.

Oren A. 2002. *Halophilic Microorganisms and their Environments*. Springer Netherlands, The Netherlands.

Oren A. 2004. Prokaryotic diversity and taxonomy: current status and future challenge. *Philos Trans R Soc Lond B* **359**: 623-638.

Oren A. 2006. The order Halobacteriales. In *The Prokaryotes A Handbook of the biology of Bacteria*, Vol 3 (ed. M Dworkin, S Falkow, E Rosenberg, HG Schlegel, E Stackebrandt), p. 1185.

Oren A. 2011. Diversity of Halophiles. In *Extremophiles Handbook*, (ed. K Horikoshi, G Antranikian, AT Bull, FT Robb, KO Stetter). Sringer, Japan.

Oren A. 2014. The family Halobacteriaceae. In *The Prokaryotes – Other Major Lineages of Bacteria and the Archaea*, (ed. E Rosenberg, EF DeLong, S Lory, E Stackebrandt, F Thompson). Springer-Verlag Berlin Heidelberg, Berlin, Germany.

Pal C, Papp B, Lercher MJ. 2005. Adaptative evolution of bacterial metabolic networks by Horizontal Gene Transfer. *Nat Genet* **37**: 1372–1375.

Papke T, Zhaxybayeva O, Feil E et al. 2007. Seaarching for species in haloarchaea. *Proc Natl Acad Sci USA* **104**(35): 14092-14097.

Patel GB, Sprott GD. 1990. Methanosaeta concilii gen. nov., sp. nov. (''*Methanothrix concilii*'') and Methanosaeta thermoacetophila nov. rev., comb. nov. *Int. J. Syst. Bacteriol.* 40: 79–82.

Paterek JR, Smith PH. 1985. Isolation and characterization of a halophilic methanogen from great salt lake. *Appl Environ Microbiol* 50:877–881.

Paterek JR, Smith PH. 1988. *Methanohalophilus mahii* gen. nov., sp. nov., a methylotrophic halophilic methanogen. *Int J Syst Bacteriol* 38:122–123.

Paul K, Nonoh JO, Mikulski L, Brune A. 2012. "Methanoplasmatales," Thermoplasmatales-related Archaea in terminte guts and other environments, are the seventh order of methanogens. *Appl Environ Microbiol* **78**(23): 8245-8253.

Peinemann S, Muller V, Blaut M, Gottschalk G. 1988. Bioenergetics of methanogenesis from acetate by Methanosarcina barkeri. *J Bacteriol* **170**: 1369-1372.

Pester M, Schleper C, Wagner M. 2011. The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr Opin Microbiol* **14**(3): 300–306.

Pflüger K, Ehrenreich A, Salmon K et al. 2007. Identification of genes involved in salt adaptation in the archaeon Methanosarcina mazei Gö1 using genome-wide gene expression profiling. *FEMS Microbiol* **277**(79-89).
Pomper BK, Saurel O, Milon A, Vorholt JA. 2002. Generation of formate by the formyltransferase/hydrolase complex (Fhc) from Methylobacterium extorquens AM1. *FEBS Lett* **523**(1-3): 133-137.
Pomper BK, Vorholt JA, Chistoserdova L et al. 1999. A methyl tetrahydromethanopterin cyclohydrolase and a methenyl tetrahydrofolate cyclohydrolase in Methanobacerium extorquens AM1. *Eur J Biochem* **1**(2): 475-480.
Popa O, Dagan T. 2011. Trends and barriers to lateral gene transfer in prokaryotes. *Curr Opin Microbiol* **14**: 1-9.
Popa O, Hazkani-Covo E, Landan G et al. 2011. Directed networks reveals genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Res* **21**: 99-609.
Prat L, Heinemann IU, Aerni HR et al. 2012. Carbon source-dependent expansion of the genetic code in bacteria. *Proc Natl Acad Sci USA* **109**: 21070-21075.
Price CE, Driessen JM. 2010. Biogenesis of membrane bound respiratory complexes in *Escherichia coli*. *Biochim Biophys Acta* **1803**: 748-766.
Price MN, Arkin AP, Alm EJ. 2006. The life-cycle of operons. *PLoS Genet* **2**(6): e96.
Price MN, Dehal PS, Arkin AP. 2008. Horizontal gene transfer and the evolution of transcriptional regulation in Escherichia coli. *Genome Biol* **9**: R4.
Price MN, Huang KH, Arkin AP, Alm EJ. 2005. Operon formation is driven by co-regulation and not by horizontal gene transfer. *Genome Res* **15**: 809-819.
Puigbo P, Wolf YI, Koonin EV. 2010. Teh tree and net components of prokaryote genome evolution. *Genome Biol Evol* **2**: 745-756.
Punta M, Coggill P, Eberhardt RY et al. 2012. The Pfam protein families database. *Nucleic Acids Res* **40**(Database issue): D290–D301.
Ragan MA, Beiko RG. 2009. Lateral genetic transfer: open issues. *Philos Trans R Soc Lond B* **364**: 2241–2251.
Ravenhall M, Skunca N, Lassalle F, Dessimoz C. 2015. inferring Horizontal Gene Transfer. *PloS Comput Biol* **11**(5): e1004095.
Rawal N, Kelkar SM, Altekar W. 1988. Alternative roues of carbohydrate metabolism in halophilic archaebacteria. *Indian J Biochem Biophys* **25**: 674-686.
Retchless AC, Lawrence JG. 2007. Temporal fragmentation of species in bacteria. *Science* **317**: 1093-1096.
Reysenbach AL, Voytek M, Mancinelli R. 2001. *Thermophiles: Biodiversity, Ecology, and Evolution*. Springer US, New York, United Sates of America.
Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics : TIG* **16**(6): 276-277.
Richardson DJ. 2000. Bacterial respiration: a flexible process fo a changing environment. *Microbiology* **146**: 551-571.
Rinke C, Schwientek P, Sczyrba A et al. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431-437.
Roeβler M, Müller V. 2001. Osmoadaptation in bacteria and archaea: common principles and differences. *Environ Microbiol* **3**(12): 743-754.
Rohlin L, Gunsalus RP. 2010. Carbon-dependent controlelectron transfer and central carbon pathway genes for methane biosynthesis in the Archaean, Methanosarcina acetivorans.
Rosenberg E, DeLong EF, Lory S et al. 2014. *The Prokaryotes: Other Major Lineages of Bacteria and the Archaea*. Springer-Verlag Berlin Heidelberg, Berlin, Germany.
Rother M, Krzycki JA. 2010. Selenocysteine, Pyrrolysine, and the unique energy metabolism of methanogenic archaea. *Archaea* **2010**(453642): 1-14.
Rothman DH, Fournier GP, French KL et al. 2014. Methanogenic burst in the end-Permian carbon cycle. *Proc Natl Acad Sci USA* **111**(15): 5462-5467.
Roy AW. 2009. Phylogenomics: Gene duplication, unrecognized paralogy and outgroup choice. *PLOS One* **4**(2): e4568.
San Millan A, Toll-Riera M, MacLean RC. 2015. Interactions between horizontally acquired genes create a fitness cost in Pseudomonas aeruginosa. *Nat Comms* **6**(6845): 8.
Schlegel HG, Jannasch HW. 2006. Prokaryotes and Their Habitats. In *The Prokaryotes*, Vol 1 (ed. M Dworkin, S Falkow, E Rosenberg, KH Schleifer, E Stackebrandt), pp. 137–184. Springer New York, New York, United States of America.
Schlegel K, Leone V, Faraldo-Gómez JD, Müller V. 2012a. Promiscuous archaeal ATP synthase concurrently coupled to Na and H translocation. *Proc Natl Acad Sci USA* **109**(3): 947-952.
Schlegel K, Müller V. 2013. Evolution of Na+ and H+ bioenergetics in methanogenic archaea. *Biochem Soc Trans* **41**: 421-426.

Schlegel K, Welte C, Deppenmeier U, Müller V. 2012b. Electron transport during aceticlastic methanogenesis by Methanosarcina acetivorans involves a sodium-translocating Rnf complex. *FEBS Journal* **279**: 4444-4452.
Schleper C, Holz I, Janekovic D et al. 1995. A multicopy plasmid of the extremely thermophilic archaeon Sulfolobus effects its transfer to recipients by mating. *J Bacteriol* **177**: 4417–4426.
Schönknecht G, Weber PMA, Lercher MJ. 2013. Horizontal gene acquisition by eukaryotes as drivers of adaptive evolution. *Bioessays* **36**: 9-20.
Schut GJ, Boyd ES, Peters JW, Adams MWW. 2012. The modular respiration complex involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and evolutionary implications. *FEMS Microbiol Rev* **37**(2): 182-203.
Segerman B. 2012. The genetic integrity of bacterial species: the core genome and the accesory genome, two different stories. *Front Cell Infect Microbiol* **2**.
Shen P, Huang HV. 1986. Homologous recombination in Escherichia coli: dependence on subtrate length and homology. *Genetics* **112**: 441–457.
Shen Y, Buick R, Canfield DE. 2001. Isotopic evidence for microbial sulphate reduction in the early Archaean era. *Nature* **410**: 77–81.
Shi W, Moon CD, Leahy SC et al. 2014. Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res* **24**: 1517-1525.
Simankova MV, Parshina SN, Tourova TP et al. 2001. *Methanosarcina lacustris* sp. nov., a new psychrotolerant methanogenic archaeon from anoxic lake sediments. *Syst Appl Microbiol* **24**: 362-367.
Shimizu S, Upadhye R, Ishijima Y, Naganuma T. 2011. Methanosarcina horonobensis sp. nov., a methanogenic archaeon isolated from a deep subsurface Miocene formation. *Int J Syst Evol Microbiol.* 61:2503-2507.
Singh N, Kendall MM, Liu Y, Boone DR. 2005. Isolation and characterization of methylotrophic methanogens from anoxic marine sediments in Skan Bay, Alaska: description of Methanococcoides alaskense sp. nov., and emended description of Methanosarcina baltica. *Int J Syst Evol Microbiol* **55**: 2531–2538.
Smith KS, Ingram-Smith C. 2007. *Methanosaeta*, the forgotten methangen? *TRENDS Microbiol* **15**(4): 150-155.
Soares JA, Zhang L, Pitsch RL et al. 2005. The residue mass of L-Pyrrolysine in three distinct methylamine methyltransferases. *J Biol Chem* **280**: 36962-36969.
Söhngen C, Bunk B, Podstawka A et al. 2016. BacDive - The bacterial diversity Metadatabase. *Nucleic acids research* **44**(4): D581-D585.
Sorokin DY, Makarova AS, Abbas B et al. 2017. Discovery of extremely halophilic, methyl-reducing euryarchaea provides insights into the evolutionary origin of methanogenesis. *Nat Microbiol* **2**: 17081.
Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. *Nat Rev Gen* **16**: 472–482.
Sowers KR, Baron SF, Ferry JG. 1984. *Methanosarcina acetivorans* sp. nov., an acetotrophic methane-producing bacterium isolated from marine sediments. *Appl Environ Microbiol* **47**: 971-978.
Sowers KR, Gunsalus RP. 1995. Halotolerance in *Methanosarcina* spp.: Role of N-acetyl-β-lysine, α-glutamate, glycine betaine, and K+ as compatible solutes for osmotic adaptation. *Appl Environ Microbiol* **61**(12): 4382-4388.
Spang A, Saw JS, Jørgensen SL et al. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**: 173-179.
Spanheimer R, Müller V. 2008. The molecular basis of salt adaptation in *Methanosarcina mazei* Gö1. *Arch Microbiol* **190**: 271-279.
Sprenger WW, van Belzen MC, Rosenberg J, Hackstein JHP, Keltjens JT. 2000. *Methanomicrococcus blatticola* gen. nov., sp. nov., a methanol- and methylamine-reducing methanogen from the hindgut of the cockroach Periplaneta americana. *Int J Syst Evol Microbiol* 50:1989–1999.
Spring S, Scheuner C, Lapidus A et al. 2010. The genome sequence of *Methanohalophilus mahii* SLPT reveals difference in the energy metabolism among members of the *Methanosarcinaceae* inhabiting frewshwater and saline environments. *Archaea* **2010**: 1-16.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)* **22**(21): 2688-2690.
Steiglmeier M, Alves RJE, Schleper C. 2014. The phylum Thaumarchaeota. In *The Prokaryotes: Other Major Lineages of Bacteria and the Archaea*, Vol 4 (ed. E Rosenberg, EF DeLong, S Lory, E Stackebrandt, F Thompson), pp. 347 - 362. Springer-Verlag Berlin Heidelberg, Berlin, Germany.
Stewart CN, Halfhill MD, Warwick SI. 2003. Transgene introgression from genetically modified crops to their wild relatives. *Nat Rev Genet* **4**: 806-8017.
Struck TH. 2013. The impact of paralogy on phylogenomics studies - A case study on annelid relationships. *PLoS ONE* **8**(5): e62892.
Stupperich E, Hammel KE, Fuchs G, Thauer RK. 1983. Carbon monoxide fixation into the carboxyl group of acetyl coenzyme A during autotrophic growth of Methanobacterium. *FEBS Lett* **152**(1): 21-23.

Swithers K, Soucy MS, Gogarten JP. 2012. The role of reticulate evolution in creating innovations and complexity. *Int J Evol Biol* **2012**: 10.

Takai K, Gamo T, Tsunogai U et al. 2004. Geochemical and microbiological evidence for a hydrogen-based, hyperthermophilic subsurface lithoautotrophic microbial ecosystem (HyperSLiME) beneath an active deep-sea hydrothermal field. *Extremophiles* **8**(4): 269-282.

Tatum EL, Lederberg J. 1947. Gene recombination in the bacterium Escherichia coli. *J Bacteriol* **53**: 673–684.

Tatusov RL. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**: 41.

Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* **278**(5338): 631-637.

Techtmann SM, Lebedinsky AV, Colman AS et al. 2012. Evidence for horizontal gene transfer of anaerobic carbon monoxide dehydrogenases. *Front Microbiol* **3**(132): 1-16.

Thauer RK. 1998. Biochemistry of methanogens: a tribute to Marjory Stephenson. 1998 Marjory Stephenson Prize Lecture. *Microbiology* **144**: 2377-2406.

Thauer RK, Kaster AK, Seedorf H et al. 2008. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Micro* **6**: 579-591.

Thomas MC, Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer bewteen Bacteria. *Nat Rev Micro* **3**: 711-721.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forgeeukaryotic chromosomes. *Nature Rev Genet* **5**: 123–135.

Ueno Y, Yamada K, Yoshida N et al. 2006. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **440**: 516-519.

Vanwonterghem I, Evans PN, Parks DH et al. 2016. Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota *Nat Microbiol* **1**: 16170.

Vascu K, Nagaraja V. 2013. Diverse functions of Restriction-Modification systems in addition to cellular defense. *Microbiol Mol Biol Rev* **77**(1): 53-72.

Vazquez-Boland JA, Dominguez-Bernal G, Gonzalez-Zorn B et al. 2001. Pathogenicity islands and virulence evolution in Listeria. *Microbes Infect* **3**(7): 571-584.

Vogan AA, Higgs PG. 2011. Teh advantages and disadvatages of horizontal gene transfer and the emergence of the first species. *Biol Direct* **6**: 14.

von Klein D, Arab H, Völker H, Thomm M. 2002. *Methanosarcina baltica*, sp. nov., a novel methanogen isolated from the Gotland Deep of the Baltic Sea. *Extremophiles* **6**: 103-110.

Vorholt JA, Chistoserdova L, M. SS et al. 1999. Distribution of tetrahydromethanopterin-Dependent enzymes in methylotrophic bacteria and phylogeny of methenyl tetrahydromethanopterin cyclohydrolases *J Bacteriol* **181**(18): 5750-5757.

Walter ME, Ortiz A, Sondgeroth C et al. 2016. High-throughput mutation, selection, and phenotype screening of mutant methanogenic archaea. *J Microbiol Methods* **131**: 113-121.

Wang F-P, Zhang Y, Chen Y et al. 2014. Methanotrophic archaea possessing diverging methane-oxidizing and electron-transporting pathways. *ISME J* **8**: 1069-1078.

Wang M, Tomb JF, Ferry JG. 2011. Electron transport in aceate-grown *Methanosarcina acetivorans*. *BMC Microbiol* **11**: 165.

Wagner D, Schirmack J, Ganzert L, Morozova D, Mangelsdorf K. 2013. *Methanosarcina soligelidi* sp. nov., a desiccation- and freeze-thaw-resistant methanogenic archaeon from a Siberian permafrost-affected soil. *Int J Syst Evol Microbiol* 63:2986–2991.

Welander PV, Metcalf WW. 2005. Loss of the mtr operon in Methanosarcina blocks growth on methanol, but not methanogenesis, and reveals an unkown methanogenic pathway. *Proc Natl Acad Sci USA* **102**(30): 10664-10669.

Wellner A, Lurie MN, Gophna U. 2007. Complexity, connectivity, and duplicability as barriers to lateral gene transfer. *Genome Biol* **8**: R156.

Welte C, Deppenmeier U. 2011. Membrane-bound electron tranport in Methanosaeta thermophila. *J Bacteriol* **193**(11): 2868-2870.

Welte C, Deppenmeier U. 2013. Bioenergetics and anaerobic respiratory chains of aceticlastic methanogens. *Biochim Biophys Acta* **1837**(7): 1130-1147.

Welte C, Kallnik V, Grapp M et al. 2010. Function of Ech hydrogenase in ferredoxin-dependent, membrane-bound electron transport in Methanosarcina mazei. *J Bacteriol* **192**(3): 674-678.

Welte C, Kröninger L, Deppenmeier U. 2014. Experimental evidence of an acetate transporter protein and characterization of acetate activation in aceticlastic methanogenesis of Methanosarcina mazei. *FEMS Microbiol Lett* **359**: 147-153.

Wiedenbeck J, Cohan FM. 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev* **35**(5): 957-576.

Wiedenheft B, Sternberg SH, Doudna JA. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**: 331-338.

Wilharm T, Zhilina TN, Hummel P. 1991. DNA-DNA hybridization of methylotrophic halophilic methanogenic bacteria and transfer of *Methanococcus halophilus* VP to the genus *Methanohalophilus* as *Methanohalophilus halophilus* comb. nov. *Int J Syst Bacteriol* 41:558–562.

Williams D, Gogarten JP, Papke T. 2012. Quantifying homologous replacements of loci between haloarchaeal species. *Genome Biol Evol* **4**(12): 1223-1224.

Williams TA, Foster PG, Cox CJ, Embley TM. 2013. An archaeal origin of eukaryotes supports only two promary domains of life. *Nature* **504**: 231–236.

Woese C. 2002. On the evolution of cell. *Proc Natl Acad Sci USA* **99**: 8742-8747.

Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* **87**: 4576–4579.

Wolf YI, Koonin EV. 2013. Genome reduction as the dominant mode of evolution. *Bioessays* **35**(9): 829-837.

Wood GE, Haydock AK, Leigh JA. 2003. Function and regulation of the formate dehydrogenase genes of the methaongenic archaeon *Methanococcus maripaludis*. *J Bacteriol* **185**: 2548-2554.

Woodson JD, Zayas CL, Escalante-Semerana C. 2003. A new pathway for salvaging the coenzyme B12 precursor cobinamide in archaea requires cobinamide-phosphate synthase (CbiB) enzyme activity. *J Bacteriol* **185**(24): 7193–7201.

Worm P, Koehorst JJ, Visser M et al. 2014. A genomic view on syntrophic versus non-syntrophic lifestyle in anaerobic fatty acid degrading communities. *Biochim Biophys Acta* **1837**(7): 2004-2016.

Worth LJ, Clark S, Radman M, Modrich P. 1994. Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. *Proc Natl Acad Sci USA* **91**: 3238–3241.

Xie G, Booner CA, Brettin T et al. 2003. Lateral gene transfer and ancient paralogy of operons containing redundant copies of tryptophan-pathway genes in *Xylella* species and in heterocystous cyanobacteria. *Genome Biol* **4**(R14): 18.

Yap WH, Zhang Z, Wang Y. 1999. Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of the entire rRNA operon. *J Bacteriol* **181**(7): 5201-5209.

Yin T, Cook D, Lawrence M. 2012. ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biol* **13**(8): R77.

Youngblut ND, Wirth JS, Henriksen JR et al. 2015. Genomic and phenotypic differentiation among Methanosarcina mazei populaiton from Columbia river sediment. *ISME J* **9**: 2191-2205.

Yu IK, Kawamura F. 1987. *Halomethanococcus doii* gen. nov., sp. nov.: an obligately halophilic methanogenic bacterium from solar salt ponds. *J Gen Appl Microbiol* 33:303–310.

Yutin N, Puigbo P, Koonin EV, Wolf YI. 2012. Phylogenomics of prokaryotic ribosomal proteins. *PLoS One* **7**(5).

Zerulla K, Soppa J. 2014. Polyploidy in haloarchaea: advantages for growth and survival. *Front Microbiol* **5**: 274.

Zhang G, Jiang N, Liu X, Dong X. 2008. Methanogenesis from methanol at low temperatures by a novel psychrophilic methanogen, "*Methanolobus psychrophilus*" sp. nov., prevalent in Zoige wetland of the Tibetan plateau. *Appl Environ Microbiol.* 74(19):6114-20.

Zhaxybayeva O. 2009. Detection and quantitative assessment of horizontal gene transfer. In *Horizontal Gene Transfer: Genomes in flux*, Vol 532 (ed. MB Gogarten, JP Gogarten, L Olendzenzki), pp. 196-213. Humana Press, New York, USA.

Zhilina TN, Zavarzin GA. 1987. *Methanohalobium evestigatus* n. gen., n. sp. The extremely halophilic methanogenic Archaebacterium. *Dokl Akad Nauk SSSR* 293:464–468 (in Russian).

Zhilina TN, Svetlichnaya TP. 1989. Ultrafine structure of *Methanohalobium evestigatum*, an extremely halophilic methanogenic bacterium. *Microbiology* (Russ) 58:248–253. (Translated from Mikrobiologiya 58:312–318).

Zinder SH, Sowers KR, Ferry JG. 1985. *Methanosarcina thermophila* sp. nov., a thermophilic acetotrophic, methane-producing bacterium. *Int J Syst Bacteriol* 35:522–523.

Zinder SH, Mah RA. 1979. Isolation and characterization of a thermophilic strain of Methanosarcina unable to use H_2-CO_2 for methanogenesis. *Appl Environ Microbiol* 38:996–1008.

APENDIX I

SUPPLEMENTARY MATERIAL: CHAPTER 5

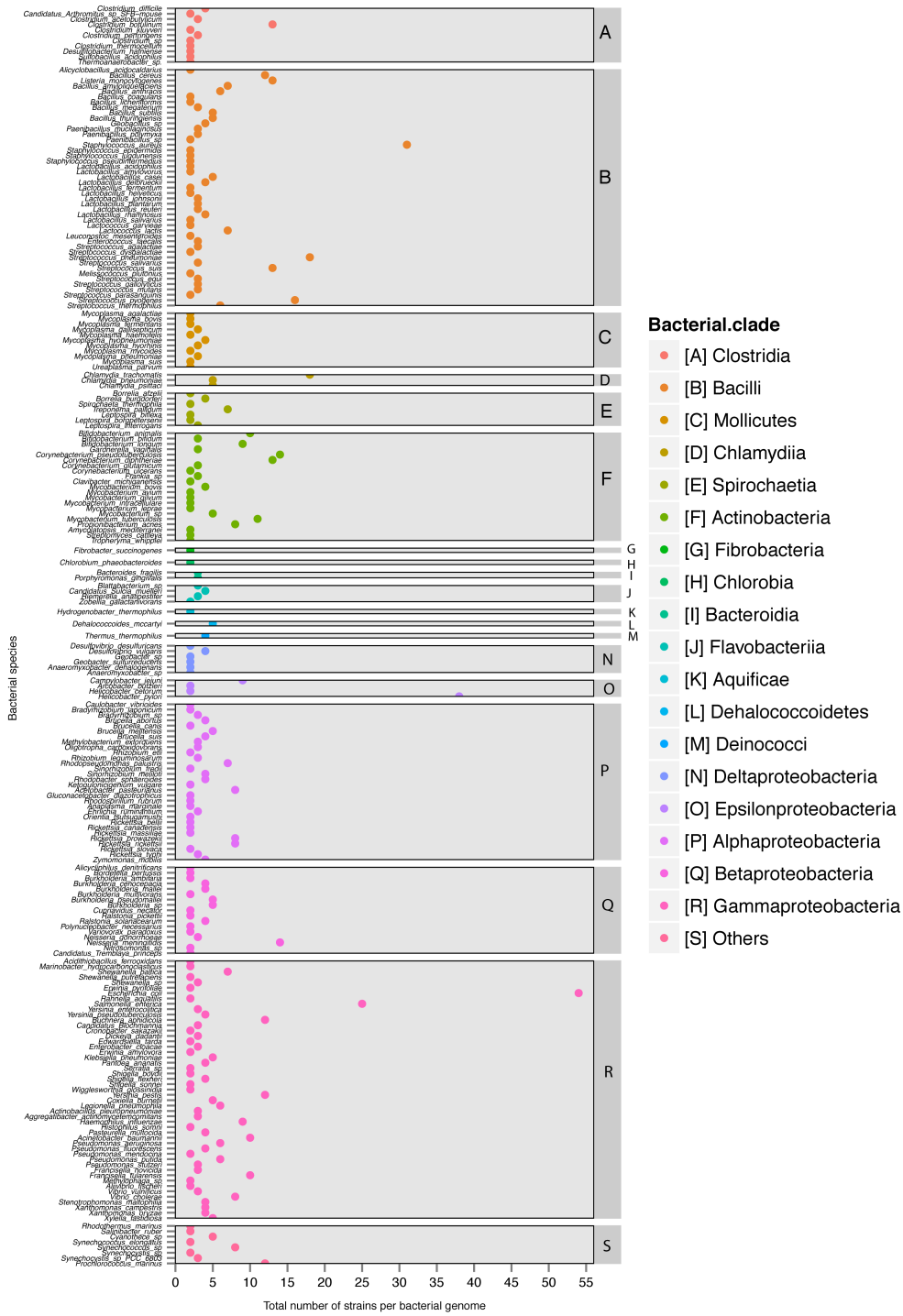


Figure SF1. Bacteria genomes strains. 1,013 strains were identified among 223 species in 19 bacterial groups. 6,247 protein families contained ≥2 sequences belonging to the same species were identified. For each family, pairwise global alignments were performed using the Needleman-Wunsch algorithm. Next, proteins from the same species in a family showing similarities ≥90% were removed. The following species have the highest (≥16 and ≤54) number of strains in the database: *Streptococcus pyogenes*, *S. pneumoniea*, *Chlamydia trachomatis*, *Salmonella enterica*, *Staphylococcus aureus*, *Helicobacteri pylori*, and *Escherichia coli*.

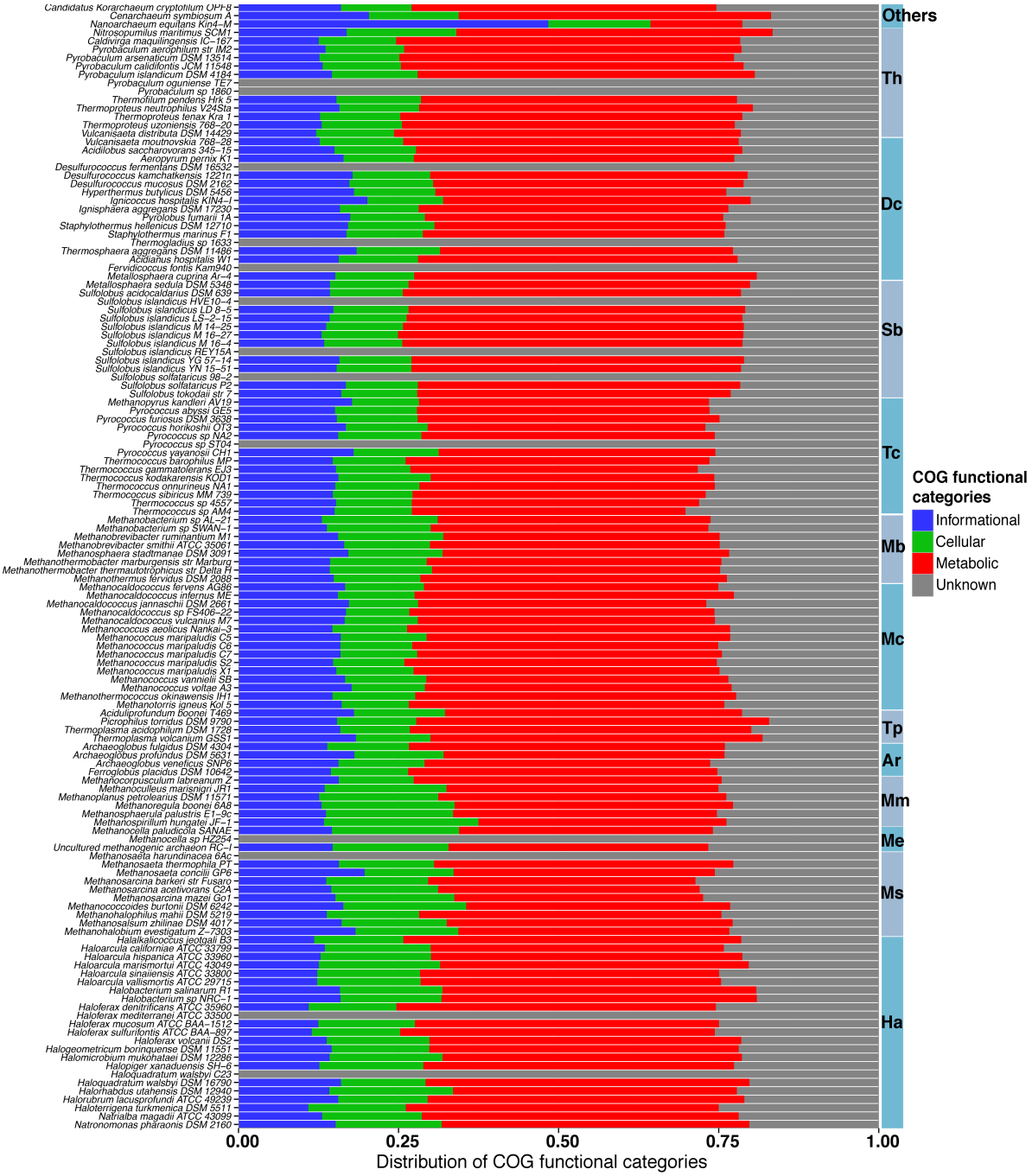


Figure SF2. COG functional categories for bacterial-like genes in each archaeal genome. Among the 134 archaeal genomes, 267,568 proteins were annotated using the Cluster of Orthologous Genes (COG) from the NCBI database. For each genome, four COG functional categories are shown: 1) Informational, 2) Cellular, 3) Metabolism, and 4) Predicted or unknown functions. Percentage of the bacterial-like genes contained in archaeal genomes covers from 36% to 65%. For 13 species predicted or unknown annotations were provided. Archaeal species are sorted based on the phylogenetic reference tree described in Figure 6 (Chapter 6), and grouped in 13 groups. Archaeal groups from the top to the bottom are as follow: **Others**, Thauma (2), Kora (1), Nanoarchaea (1); **Th**, Thermoproteales (13); **Dc**, Desulfurococcales (13); **Sb**, Sulfolobales (17); **Tc**, Thermococcales (14); **Mb**, Methanobacteriales (8); **Mc**, Methanococcales (15); **Tp**, Thermoplasmatales (4); **Ar**, Archaeogobales (4); **Mm**, Methanomicrobiales (6); **Me**, Methanocellales (3); **Ms**, Methanosarcinales (10); **Ha**, Haloarchaea (23).

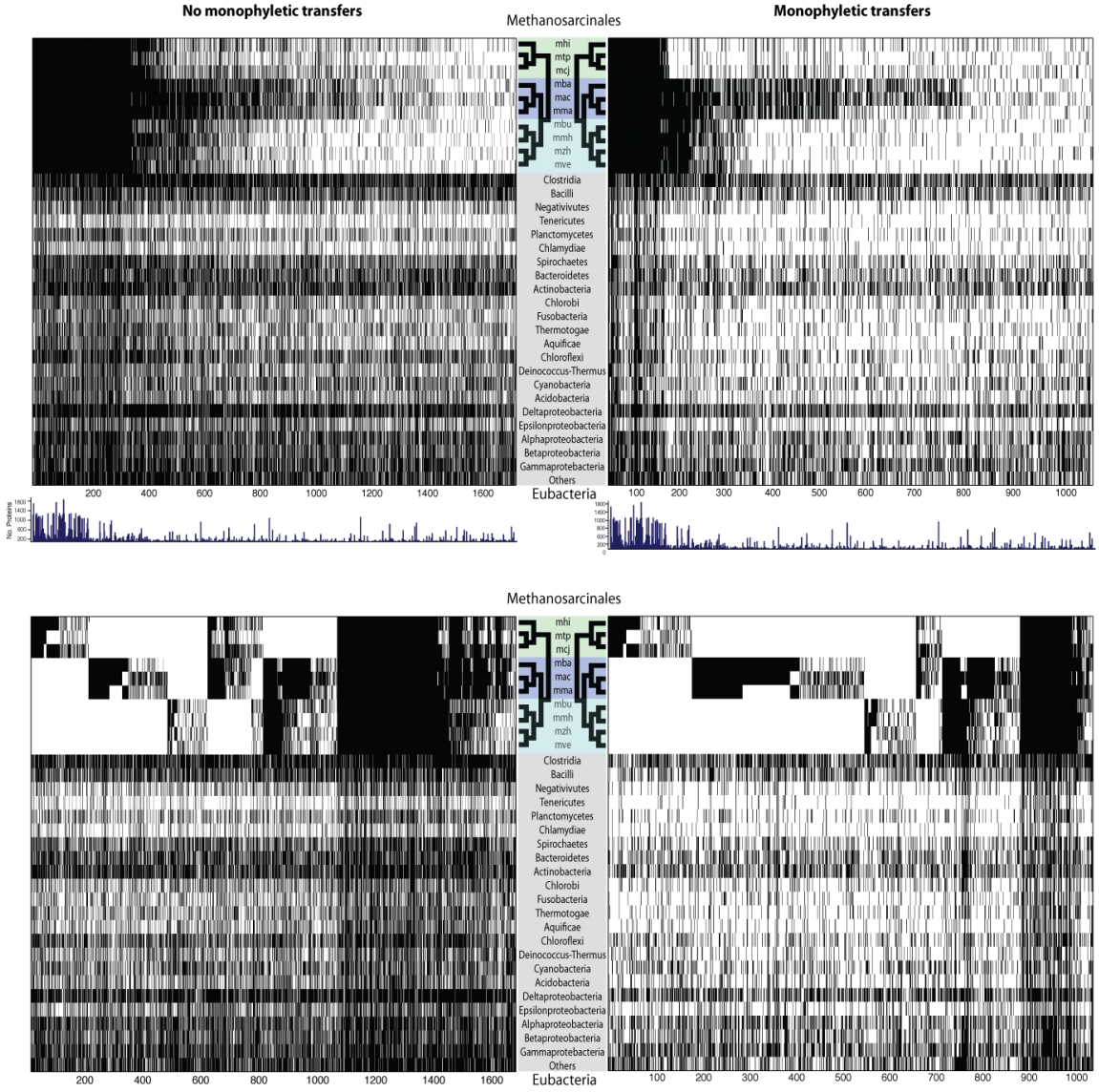


Figure SF3. Gene distributions of HGTs in across all Methanosarcinales members. Methanosarcinales members are shown in at top of each figure (reference topology for Methanosarcinales in Figure 5.1). The three main branches correspond to taxonomic divisions (colored sections): *Methanosaeta* (mhi, mtp, mcj), *Methanosarcina* (mba, mac, mma) and a group of 4 close related species (mbu, mmh, mzh, mve) to *Methanosarcina*. *Species code*: ***Methanosaeta species***, mhi, *M. harundinacea*; mtp, *M. thermophile*; mcj, *M. concilii*; mba; ***Methanosarcina species***, *M. barkeri* (*Fusaro*); mac, *M. acetivorans*; mma, *M. mazei*; ***Others subgroup***, mbu, *M. burtonii*; mmh, *M. mahii*; mzh, M. zhilinae; mve, *M. evestigatum*. 23 Bacterial groups are displayed below to the Methanosarcinales tree. Ticks indicate the presence (black) or absence (white) of the gene in the corresponding either archaeal (Methanosarcinales) and bacterial (within the group) genome. *Left panel*: HGTs assigned to 1,770 non-monophyletic cases. Transfers are distributed through Methanosarcinales members (from whole to a patchy presence in genome); the corresponding presence in other bacterial groups is at the bottom (upper panel). Distribution of HGTs through the seven created subgroups of Methanosarcinales; the presence of the gene in any bacterial group is shown below (lower panel). *Right panel*: Distribution of 1,045 imports in all Methanosarcinales genomes. Imports distributions are from whole to a patchy presence in genome (upper panel); and displayed in each subgroup of Methanosarcinales (lower panel). Protein family size is shown for each transfer category (upper panels at the bottom).

Table ST1. Bacterial gene donors involved with the Methanosarcinales imports.

Group	Bacterial donor	Frequency	Group	Bacterial donor	Frequency	Group	Bacterial donor	Frequency	Group	Bacterial donor	Frequency	Group	Bacterial donor	Frequency
Methanosaeta	Thermodesulfobacteria	3	Methanosarcina	Thermotogae	5	Msa-Othrs	Thermotogae	1	Others Ms	Thermotogae	2	All	Thermotogae	4
	Actinobacteria	14		Acidobacteriia	1		Acidobacteriia	1		Actinobacteria	2		Actinobacteria	6
	Alphaproteobacteria	4		Actinobacteria	10		Actinobacteria	2		Alphaproteobacteria	2		Alphaproteobacteria	9
	Anaerolineae	1		Alphaproteobacteria	3		Alphaproteobacteria	1		Bacilli	5		Aquificae	1
	Bacilli	1		Aquificae	4		Aquificae	1		Betaproteobacteria	2		Bacilli	1
	Chlamydiia	2		Bacilli	8		Bacilli	1		Chlorobia	1		Caldisericia	2
	Chlorobia	1		Bacteroidia	1		Bacteroidia	1		Clostridia	9		Chlamydiia	1
	Chloroflexi	3		Betaproteobacteria	5		Clostridia	21		Cytophagia	1		Chlorobia	2
	Clostridia	13		Caldilineae	1		Deferribacteres	1		Deferribacteres	1		Chloroflexi	2
	Dehalococcoidetes	1		Caldisericia	2		Deltaproteobacteria	15		Deinococci	1		Clostridia	15
	Deinococci	2		Chlamydiia	3		Epsilonproteobacteria	1		Deltaproteobacteria	14		Deinococci	2
	Deltaproteobacteria	21		Chlorobia	4		Flavobacteriia	1		Dictyoglomia	1		Deltaproteobacteria	7
	Epsilonproteobacteria	1		Chloroflexi	2		Fusobacteriia	1		Epsilonproteobacteria	3		Dictyoglomia	1
	Flavobacteriia	3		Chrysiogenetes	1		Gammaproteobacteria	7		Gammaproteobacteria	5		Epsilonproteobacteria	1
	Gammaproteobacteria	9		Clostridia	61		Gloeobacteria	1		Planctomycetia	1		Flavobacteriia	1
	Nitrospira	2		Cytophagia	1		Planctomycetia	2		Solibacteres	1		Fusobacteriia	1
	Sphingobacteriia	3		Deinococci	2		Sphingobacteriia	2		Sphingobacteriia	2		Gammaproteobacteria	4
	Spirochaetia	4		Deltaproteobacteria	16		Spirochaetia	4		Spirochaetia	1		Mollicutes	4
	Unclassified group	5		Dictyoglomia	1		Thermodesulfobacteria	1		Synergistia	1		Negativicutes	1
				Epsilonproteobacteria	4		Thermomicrobia	1		Thermodesulfobacteria	1		Nitrospira	1
				Flavobacteriia	7		Unclassified group	6		Thermomicrobia	1		Opitutae	1
				Fusobacteriia	5					Unclassified group	5		Planctomycetia	5
				Gammaproteobacteria	11	Mta-Msa	Actinobacteria	2					Spirochaetia	5
				Mollicutes	1		Alphaproteobacteria	1					Synergistia	3
				Nitrospira	2		Bacilli	2					Thermodesulfobacteria	3
				Planctomycetia	4		Clostridia	5					Thermomicrobia	2
				Sphingobacteriia	2		Deltaproteobacteria	5					Unclassified group	1
				Spirochaetia	8		Gammaproteobacteria	3						
				Synergistia	1		Ignavibacteria	1						
				Thermodesulfobacteria	1		Planctomycetia	1						
				Unclassified group	7	Mta-Othrs	Clostridia	1						

Bacterial gene donors were identified at the sister branch of each Methanosarcinales proteins. Bacterial donors of 518 (50% of the imports) and 52 (5% of the imports) gene trees belonging to one and ≥2 clades, respectively, were identified. In the remaining 475 gene trees a bacterial donor was not possible to be identified because proteins from Methanosarcinales or/and other archaeal clade were interleaving between bacterial branches. Most of the bacteria present in the trees belong to Firmicutes, Actinobacteria and Proteobacteria. Based on the gene distribution of imports in Figure 7.2 (upper and lower panels), a search of potential specific distribution of bacterial donors was carried out. The main results of this analysis are shown in Figure 7.4. Here, 518 import donors are presented by bacterial taxa groups, which are divided in the seven Methanosarcinales groups described in the Figure SF2-7: Methanosaeta (93), Methanosarcina (184), Others (62), Mta-Msa (20), Mta-Othr (1), Msa-Othr (72), and in All groups (86).

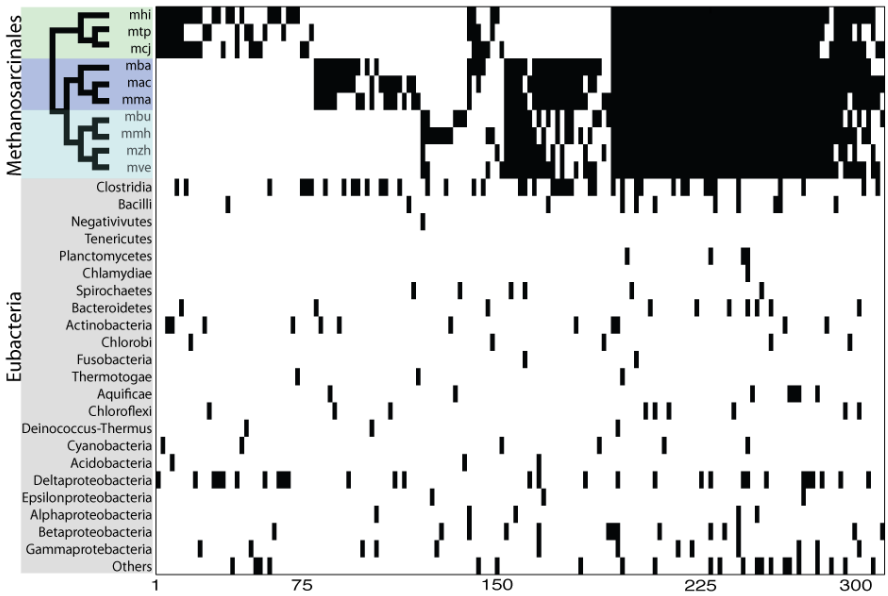


Figure SF4. Distribution of 313 exports across the seven created subgroups of Methanosarcinales. Methanosarcinales members are shown in at top of each figure (reference topology for Methanosarcinales in Figure 5.1). The three main branches correspond to taxonomic divisions (colored sections) as in figure SF3:*Methanosaeta species, Methanosarcina species*, *Others subgroup*. 23 Bacterial recipient groups are displayed below to the Methanosarcinales tree. Ticks indicate the presence (black) or absence (white) of the gene in the corresponding either archaeal (Methanosarcinales) and bacterial (within the group) genome.

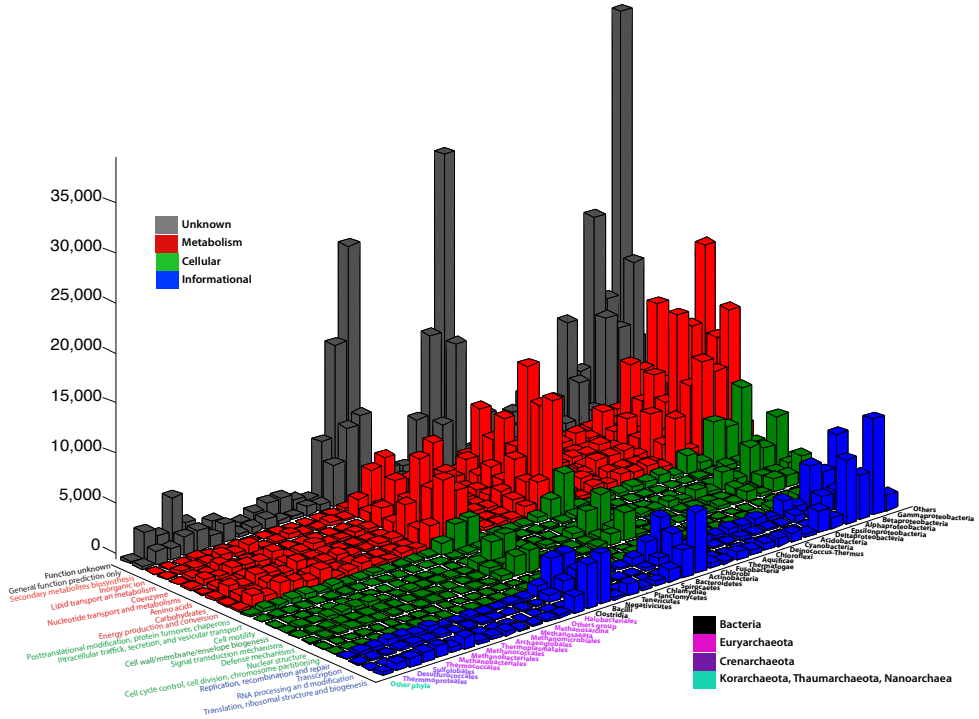


Figure SF5. Frequency of functional COG categories for prokaryotic genes in the 3,128 protein families associated to Methanosarcinales and their bacterial homologs. Proteins corresponding to single COG functional category (legends at bottom left) were countered per each prokaryotic clade (legends at bottom right). Prokaryotic proteins were grouped by their corresponding either for Archaea or Bacteria (see colored bar at bottom right). COG functional categories were divided in Metabolic (red color), Cellular (green color), Informational (blue color) and Unknown (grey color); each category is divided in different types of functions (colored labels at bottom left). Methanosarcinales member were divided in Methanosaeta, Methanosarcina and Others subgroups.

Figure SF6. Gene distribution and HGT detection of multiple proteins membrane-bound among the 134 members of Archaea [Expanded version of the Figure 7.7.1 including all members of Archaea]. The following components are shown: 10 membrane-bound complexes, 4 single permeases, and the pyrrolysine operon. **Central figure** (from left to right): Formate dehydrogenase, *fdh*; Heterodisulfide reductase, *hdr*; two F_{420} non-reducing hydrogenases, *vho* and *vht*; Rnf complex, *rnf*; Energy conservation hydrogenase, *ech*; acetate (*aceP*), tri- (mttP),di- (mtbP), and mono- (mtmP) methylamines uptake (substrates); Pyrrolysine operon, *pyl*; two ATPases (*atp* and *aha*); antiporter Na^+/H^+, *mrp*; and the Nitrogenase, *nif*. **At center:** Presence (filled circles) or absence (empty circles) of genes in the Methanosarcinales families is shown (see *Gene categories*). HGT events are described based on the tree topology (*Import*, *Export*, and *undefined transfer*), while its absence represents families without bacterial homologs (*clade specific*). Among the families, genes present in other methanogenic archaea with bacterial (black circles) or without bacterial homologs (yellow) are shown. Genes which were not found in the Methanosarcinales families, but its *presence* could be likely in other non-Methanosarcinales families, were considered as genes with *undefined protein family* (grey circles), whereas genes present in Methanosarcinales families but its *monophyly* was not tested, were designed as *undefined tree topology* (black circles). **At bottom:** Genes with ≥2 copies are shown at the bottom (pink stars). Groups: **Others**; **Th** (Thermoproteales); **Dc** (Desulfurococcales); **Sb** (Sulfolobales); **Tc** (Thermococcales); **Mb** (Methanobacteriales); **Mc** (Methanococcales); **Tp** (Thermoplasmatales); **Ar** (Archaeogobales); **Mm** (Methanomicrobiales); **Me** (Methanocellales); **Ms** (Methanosarcinales); **Ha**, Haloarchaea.

Table ST2. Distribution of genes of energy and methane metabolisms found in other protein families

where Methanosarcinales proteins are not present.

Function	Freq *	Family °	Gene	Description
Biosynthesis	1	9696	cofH	Cofactor F420 biosynthesis
	3	2085	cofH	Cofactor F420 biosynthesis
	1	8332	pylB	Pyrrolysine synthesis
Membrane-bound complexes	2	9331	frhB	Cofactor F420 reducing hydrogenase
	4	1481	frhB	Cofactor F420 reducing hydrogenase
	7	996	frhG	Cofactor F420 reducing hydrogenase
	9	935	frhA	Cofactor F420 reducing hydrogenase
	16	2717	frhD	Cofactor F420 reducing hydrogenase
	1	2493	echE	Ech hydrogenase
	6	601	echC	Ech hydrogenase
	7	5909	echF	Ech hydrogenase
	11	5770	echA	Ech hydrogenase
	18	5847	echB	Ech hydrogenase
	2	3034	mrpG	Electrochemical potential-driven transporter
	2	3867	mrpC	Electrochemical potential-driven transporter
	6	3702	mrpA	Electrochemical potential-driven transporter
	49	2841	mrpD	Electrochemical potential-driven transporter
	2	14353	vhtA	F420-nonreducing hydrogenase
	7	4497	vhtA	F420-nonreducing hydrogenase
	1	1755	fdhB	Formate dehydrogenase
	12	10704	fdhA	Formate dehydrogenase
	14	303	fdhA	Formate dehydrogenase
	20	1797	fdhB	Formate dehydrogenase
	1	1430	fpoK	F(420)H(2) dehydrogenase
	3	681	fpoL	F(420)H(2) dehydrogenase
	6	416	fpoB	F(420)H(2) dehydrogenase
	11	994	fpoC	F(420)H(2) dehydrogenase
	13	254	fpoD	F(420)H(2) dehydrogenase
	19	407	fpoH	F(420)H(2) dehydrogenase
	25	636	fpoI	F(420)H(2) dehydrogenase
	31	1268	fpoN	F(420)H(2) dehydrogenase
	3	718	hdrB	heterodisulfide reductase
	10	1335	hdrC	heterodisulfide reductase
	20	942	hdrA	heterodisulfide reductase
	49	4064	hdrD	heterodisulfide reductase
	3	9216	MtmP/MttP	Acetate Methylamines transporter
Aceticlastic methanogenesis	3	315	ACS	Acetil CoA synthetase
	1	1296	cdhE	CODH ACDS
	1	1321	cdhD	CODH ACDS
	6	2371	PPiase	Pyrophosphatase
Hydrogenotrophic methanogenesis	1	758	fwdE,fmdE	Formylmethanofuran dehydrogenase
	2	1188	mtrA2,mtrA	Formylmethanofuran dehydrogenase
	2	1906	fwdB,fmdB	Formylmethanofuran dehydrogenase
	2	991	fwdF,fmdF	Formylmethanofuran dehydrogenase
	3	1385	mtrB2,mtrB	Formylmethanofuran dehydrogenase
	14	885	fwdB,fmdB	Formylmethanofuran dehydrogenase
	21	2128	fwdH	Formylmethanofuran dehydrogenase
	1	1078	ftr	Formylmethanofuran tetrahydromethanopterin methyltransferase

* Freq: frequency of proteins detected in other families out of Methanosarcinales families (unknown protein family).

° Family: number of the protein family where the reference gene was found for the Methanosarcinales study.

Archaeal clades related to these proteins are Thermoproteales, Methanocellales, Archaeoglobales, Methanomicrobiales, Methanobacteriales, Methanococcales, Desulfuroccocales, and Sulfolobales.

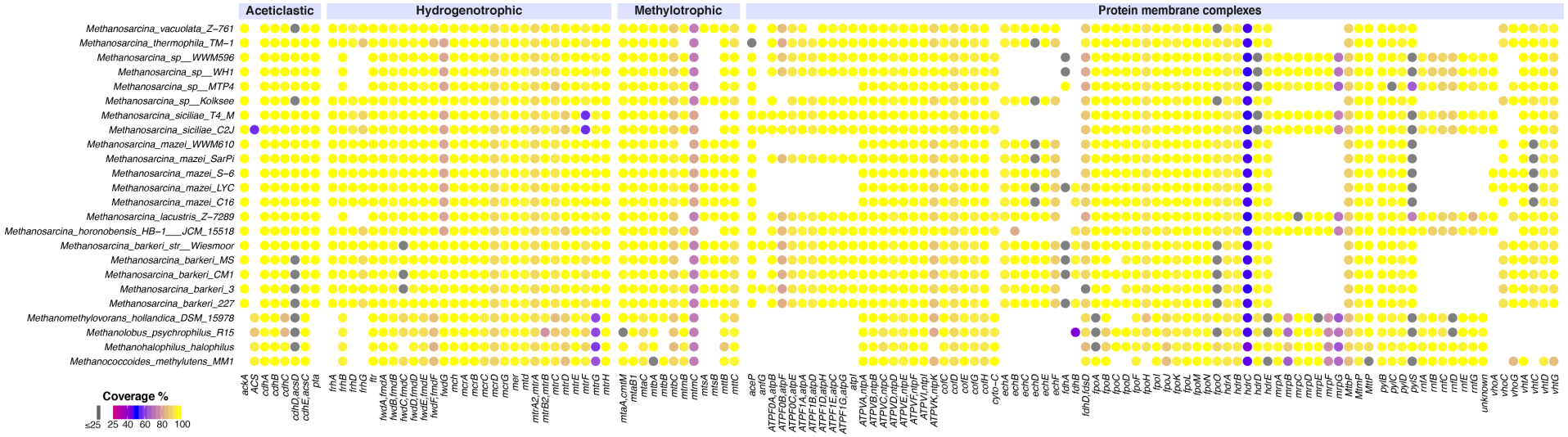


Figure SF7. Coverage percentage between homologues genes within Methanosarcinales genomes involved in methanogenesis and energy metabolism. Four major gene groups are shown; three are involved with the methanogenesis, which are the aceticlastic, methylotrophic, and hydrogenotrophic pathways, and multiple protein complexes. Methanosarcinales species are shown (left side). Genes are sorted by pathway and they are scored by the percentage of sequence coverage by color intensity (see color bar at the bottom). Absent genes are not shown (empty spaces between genes).

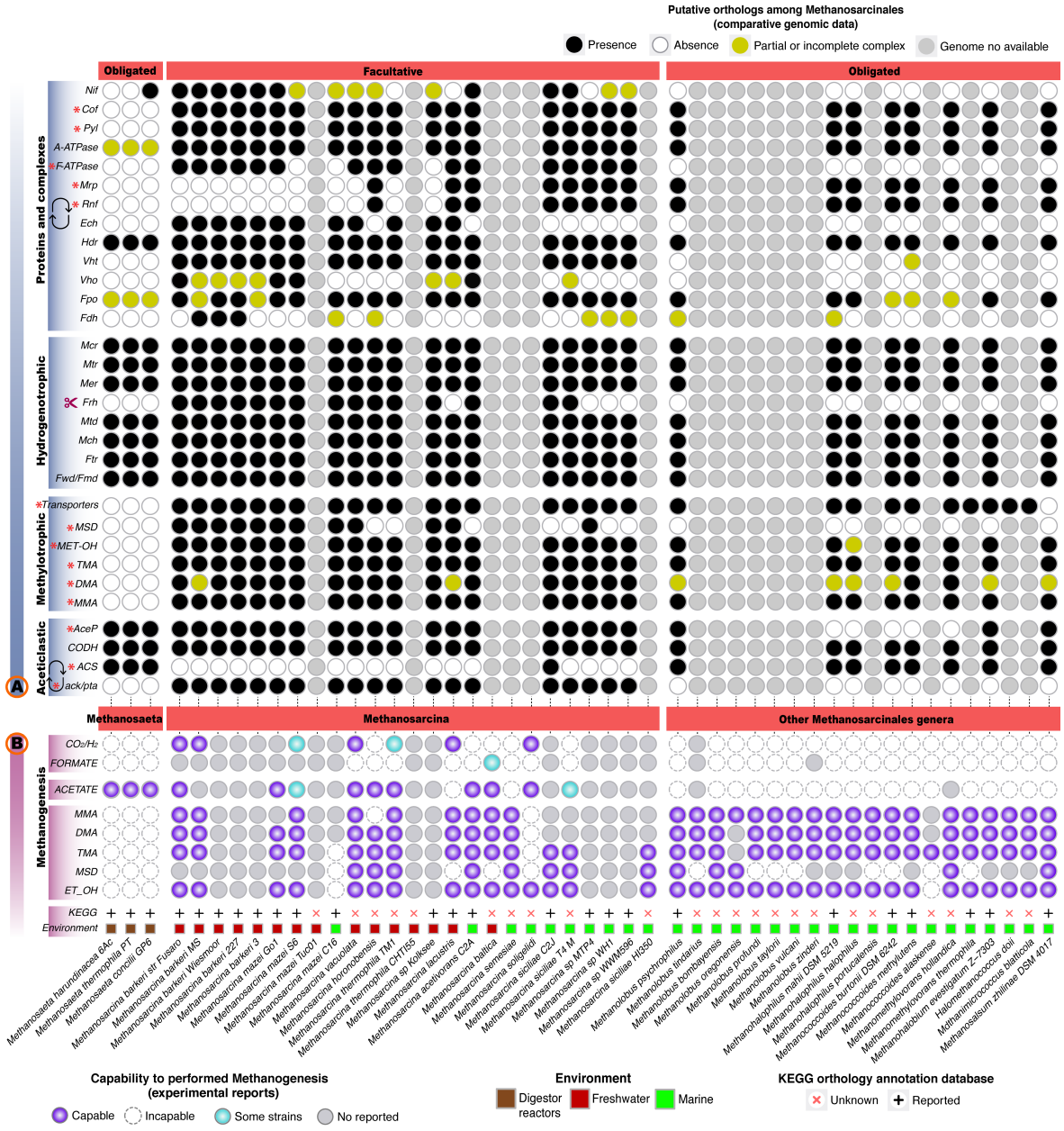


Figure SF8. Capability to perform methanogenesis: experimental and genomic data in Methanosarcinales through different pathways. As in Figure 8, the capability to perform methanogenesis using different substrates in Methanosarcinales (at the bottom) is shown in the context of genomic information (*panel A*) and experimental reports (*panel B*). **Panel A:** Orthologous genes among Methanosarcinales genomes divided in three groups: Methanosaeta, Methanosarcina, and Other Methanosarcinales genera. The presence (black circles) or absence (empty circles with grey outline) of proteins is shown for each genome in four major functions of methanogenesis and protein membranes. Incomplete complexes (yellow circles) were defined when ≥1 subunit is missed. Acquisitions (red asterisks), gene loss (purple scissor), and gene replacement (recycling arrows) are shown at the left side of each gene. **Panel B:** The 48 Methanosarcinales (species and strains) with their methanogenic capabilities (left axis) are shown. Methanosarcinales capable (purple circles) or incapable (empty circles with grey dashed outline) to grow in specific substrates are shown. Cases where some strains are capable to grow in specific conditions are shown in cyan. Species with no experimental report are displayed in grey. Methanogenic conditions are as follow: ***hydrogenotrophic*** (CO_2/H_2) and/or formate; ***aceticlastic*** (acetate); ***methylotrophic***, mono-,di-,tri-methylamine, dimethylsulfide, and methanol (MMA, DMA, TMA, DMS, MET-OH). Marine (green frame) and freshwater (red frame) lifestyles are shown at the bottom. KEGG annotations (+) were used to corroborate the gene functions; unavailable (x) annotations are shown.

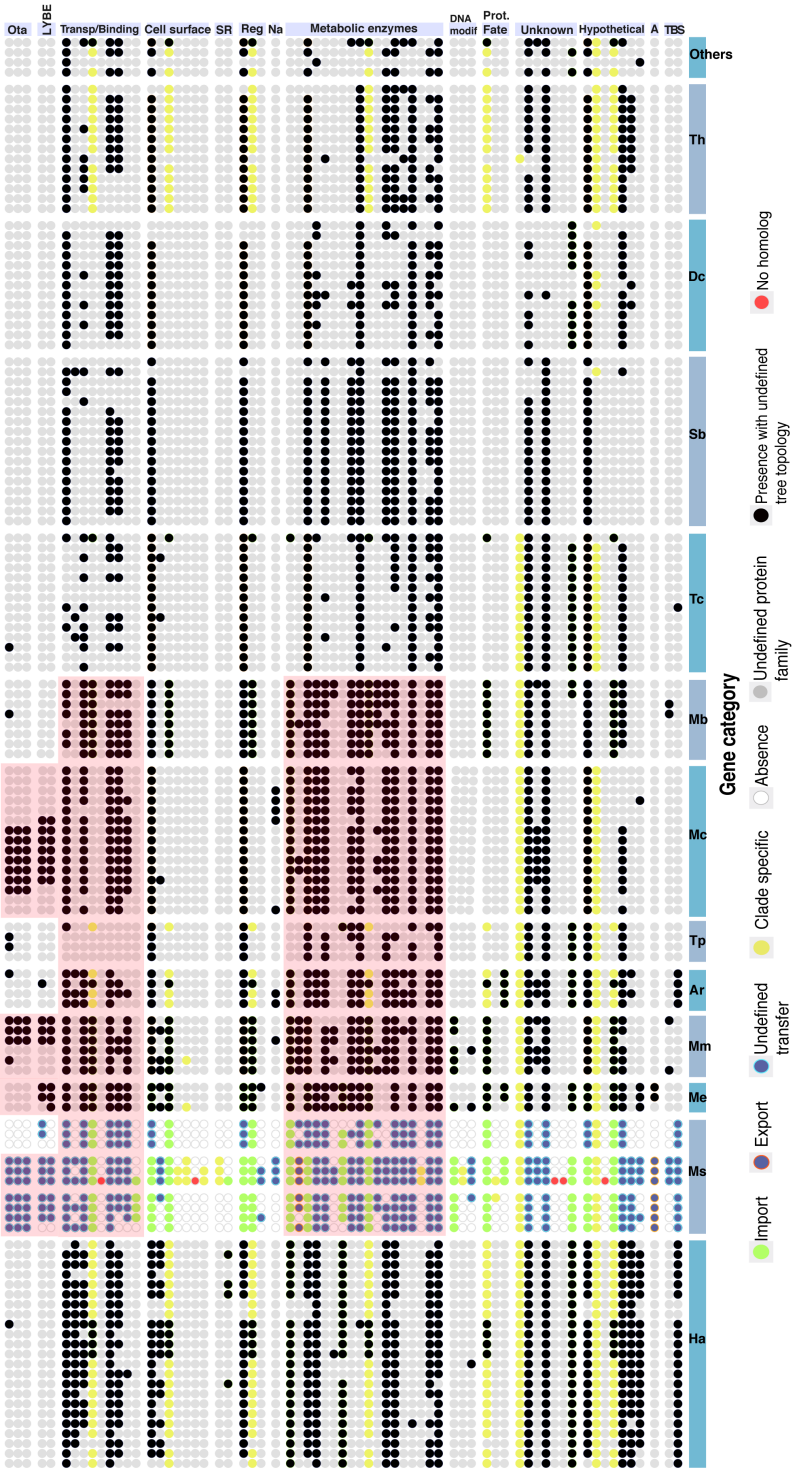


Figure SF9. (*Continue from previous page*). Gene distribution of genes involved in salt adaptation across archaea and their association to HGT events. Presence-absence pattern of 85 homologues genes related to salt adaptation in Methanosarcinales. **At the top:** Different main functions of these genes are shown; functions are equally organized like in the Figure 7.5.5 (see Table 7.5 for corresponding gene names and their specific roles). **At the right side**: Gene categories related to the topology of the phylogenetic tree are shown. **Left axis (next to the gene categories)**: Archaeal groups which are sorted as follow (from top to the bottom): **Others**; **Th (**Thermoproteales); **Dc** (Desulfurococcales); **Sb** (Sulfolobales); **Tc** (Thermococcales); **Mb** (Methanobacteriales); **Mc** (Methanococcales); **Tp** (Thermoplasmatales); **Ar** (Archaeogobales); **Mm** (Methanomicrobiales); **Me** (Methanocellales); **Ms** (Methanosarcinales); **Ha**, Haloarchaea. Methanosarcinales are sorted as in Figure 7.7.5. **Red region:** Conserved genes related to the salt-out mechanism among methanogens are shown.

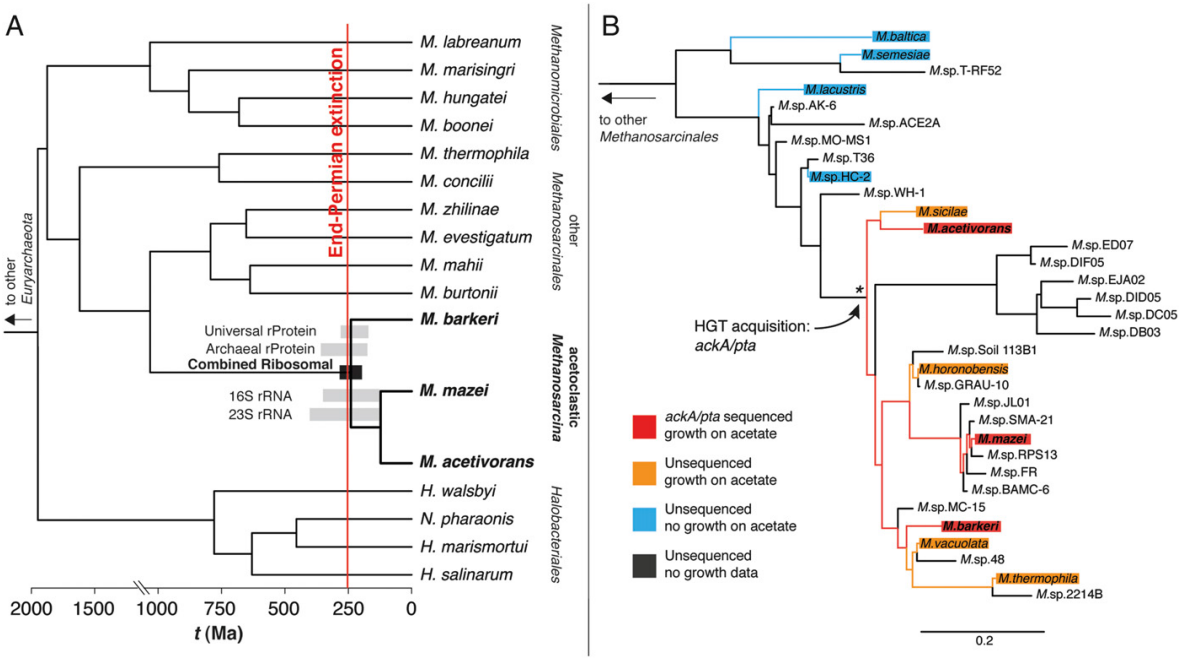


Figure SF10. Acetoclastic Methanosarcina, its association to the end-Permian event, and the acquisition of the ack/pta system. Phylogenetic relationship of methanogenic groups that link the origin of the aceticlastic pahway and the Permian extinction event (A) and Methanosarcina species (B) reported by Rothman et al. (2014). Phylogeny at section A shows the reconstruction of relaxed molecular clock chronograms based on 50 representative archaeal genomes. Shaded bars indicate the estimation of acetoclastic Methanosarcina ancestor age range. Phylogeny at section B shows Methanosarcina species that do not grow (blue labels) and do grow (red and orange labels) on acetate. The asterisk represent all descend from the ancestor containing ackA/pta genes.

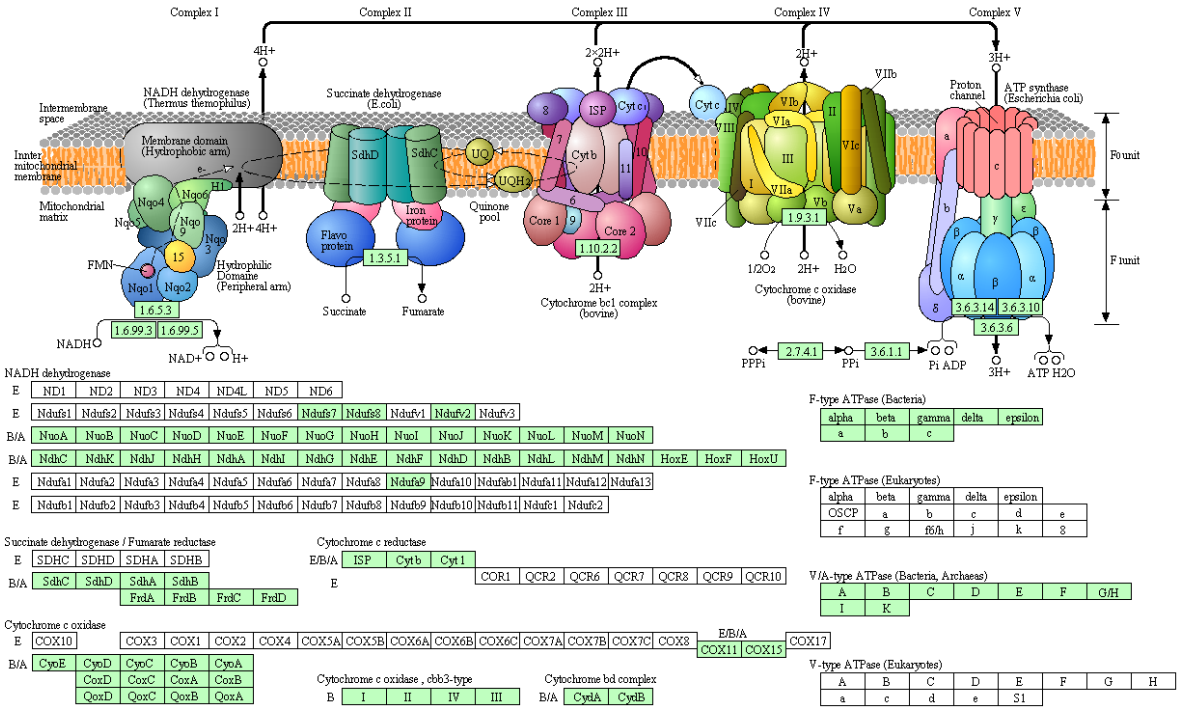


Figure SF11. Oxidative phosphorylation and membrane bound complexes in aerobic respiration. Reconstruction of metabolic pathways associated to the oxidative phosphorylation using KEGG database. **Upper panel** *(membrane-bound protein complexes)***:** Protein membrane complexes are shown at the top by their complex number, metabolic reactions and protein subunits**.** ***Lower panel*** *(proteins encompassing each membrane complex)***:** The different protein complexes (10 complexes) and their corresponding subunits (rows lined up at each complex) are shown. Identified putative orthologous genes involved in the aerobic respiratory system in Halobacteriales (through the KEGG KO orthologous system) are shown as green filled frames. The membrane-bound complexes are described as follow: **Complex I** (NADH and NAD(P)H dehydrogenase); **Complex II** (Succinate dehydrogenase and Fumarate reductase), **Complex III** (Cytochrome c reductase), **Complex IV** (Cytochrome c oxidase, cbb3-type, bd complex), and **Complex V** (F-, V/A-ATPase). Membrane complexes are organized by their presence in the three domains of life: A, Archaea; B, Bacteria; E, Eukaryota. Green color in the rows corresponds to the homolog subunits found in the prokaryotic families by the reconstruction of metabolic pathways.

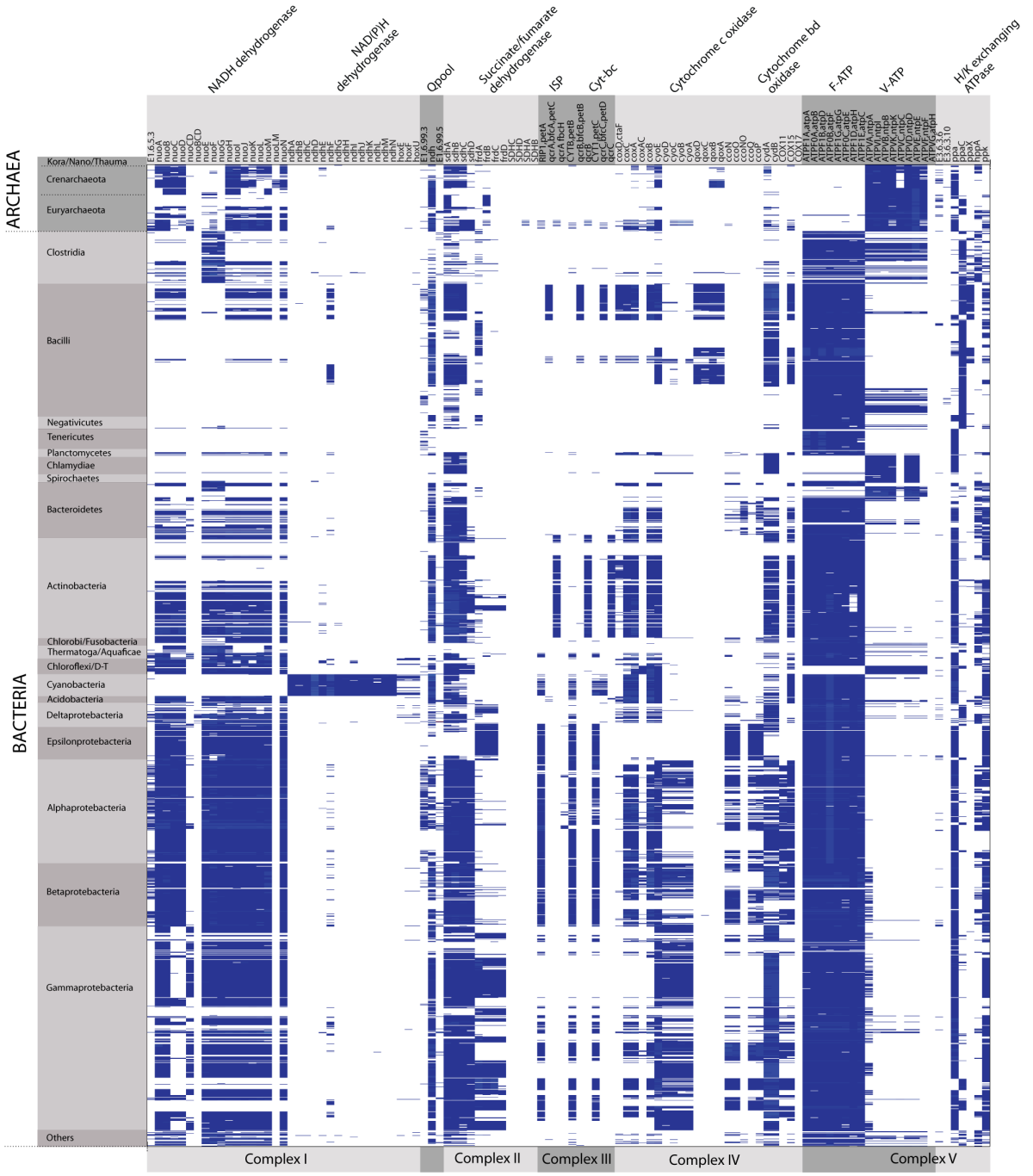


Figure SF12. Distribution of genes associated to the bacterial respiratory system across prokaryotes. Prokaryotic higher groups (left side) and the corresponding genes (at the top) to each of the five membrane-bound complex (at the bottom) are shown. Likewise, complete names of all complexes where genes were screened are show at the top. Genes involved in the respiratory system were screen across all prokaryotic genomes. As first approach, putative orthologs were identified through KEGG KO orthologs system annotation with specific parameters (e-value 10^{-10} and ≥25% global identity).

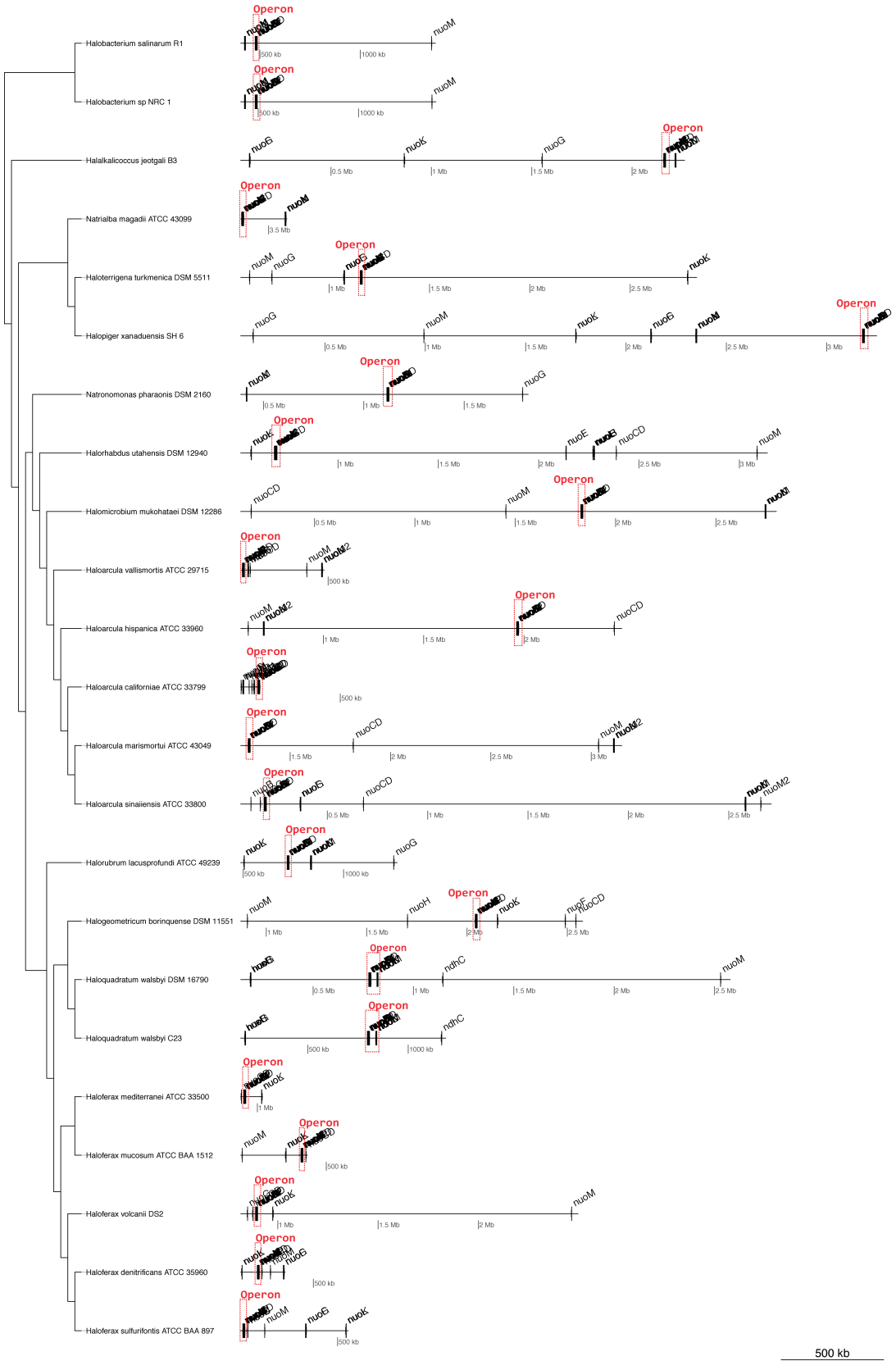


Figure SF13. Distribution genes of the respiratory system with Halobacteriales genomes: Complex I. Genes clustered as operon-like (legends and frames in red), and no clustered (without legends in red) are shown. Halobacteriales genomes are sorted by a subtree obtained from the reference tree reconstructed with concatenated universal proteins for all archaeal genomes (see Chapter 6). Genetic distances are shown in kilobases (kb).

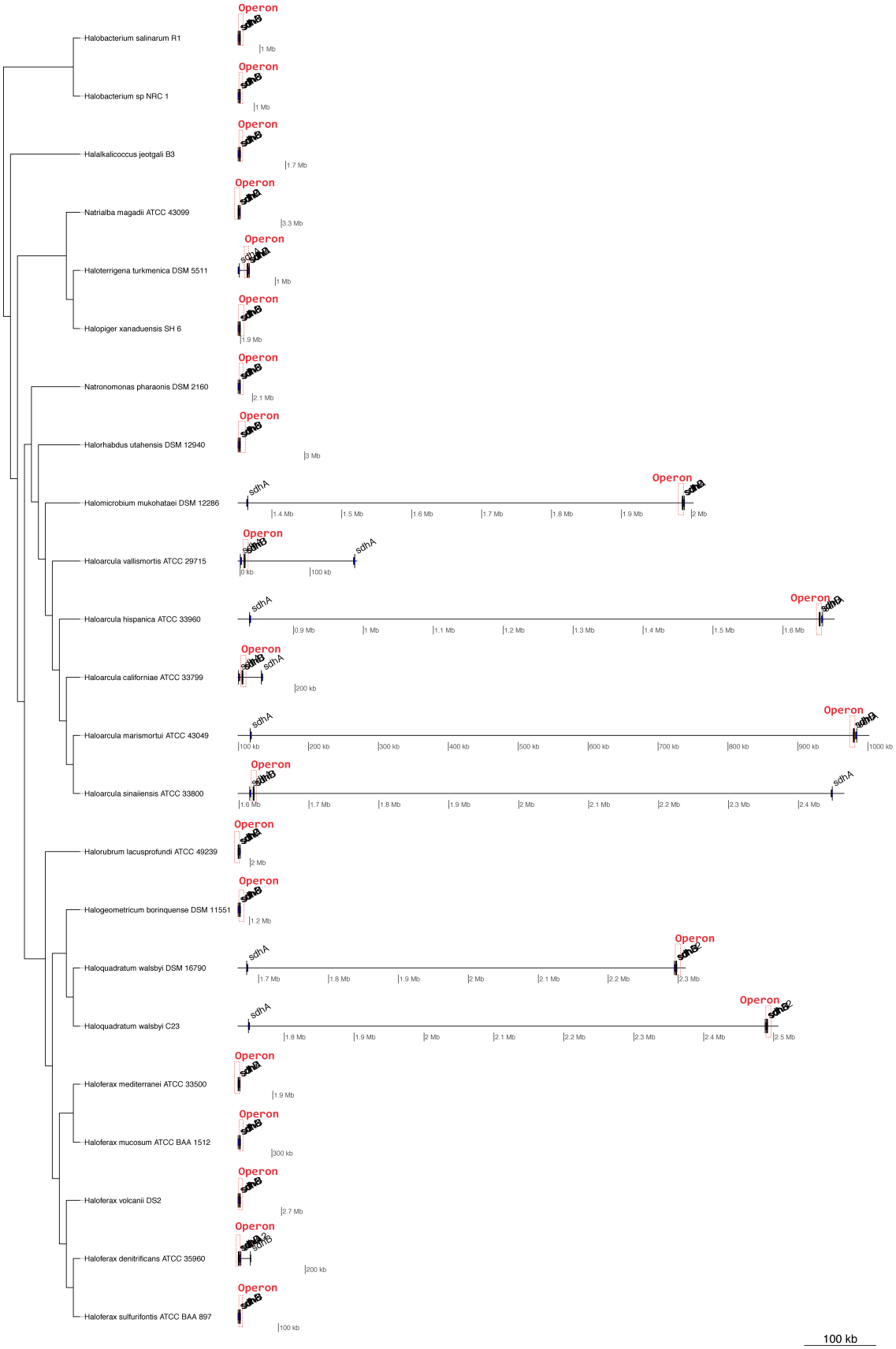


Figure SF14. Distribution genes of the respiratory system with Halobacteriales genomes: Complex II. Genes clustered as operon-like (legends and frames in red), and no clustered (without legends in red) are shown. Halobacteriales genomes are sorted as in SF3-8 (above). Genetic distances are shown in kilobases (kb).

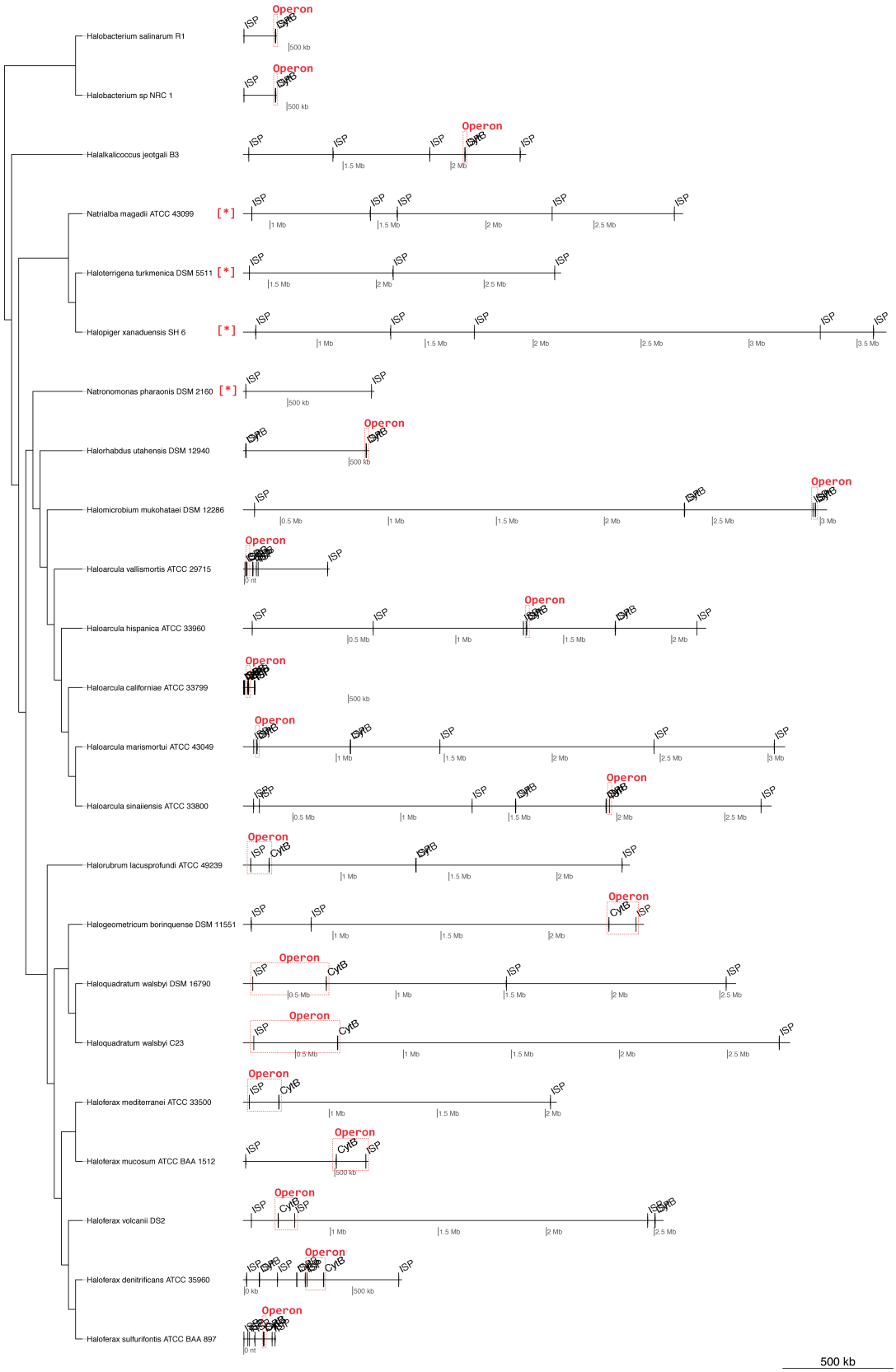


Figure SF15. Distribution genes of the respiratory system with Halobacteriales genomes: Complex III. Genes clustered as operon-like (legends and frames in red), and no clustered (without legends in red) are shown. Halobacteriales genomes are sorted as in SF3-8 (above). Asterisk (*) represents a genome with an incomplete membrane complex, where ≥1 protein is missed. Genetic distances are shown in kilobases (kb).

Figure SF16. Distribution genes of the respiratory system with Halobacteriales genomes: Complex IV. Genes clustered as operon-like (legends and frames in red), and no clustered (without legends in red) are shown. Halobacteriales genomes are sorted as in SF3-8 (above). Asterisks in red (*) represent either potential duplicated genes forming a second cluster (operon-like) or single copy genes no closely clustered composing an entire complex (asterisk with dotted frame). Genomes containing operon-like sets of genes for coxABCD (operon-A) and cyAB (operon-B) are shown. Genetic distances are shown in kilobases (kb).

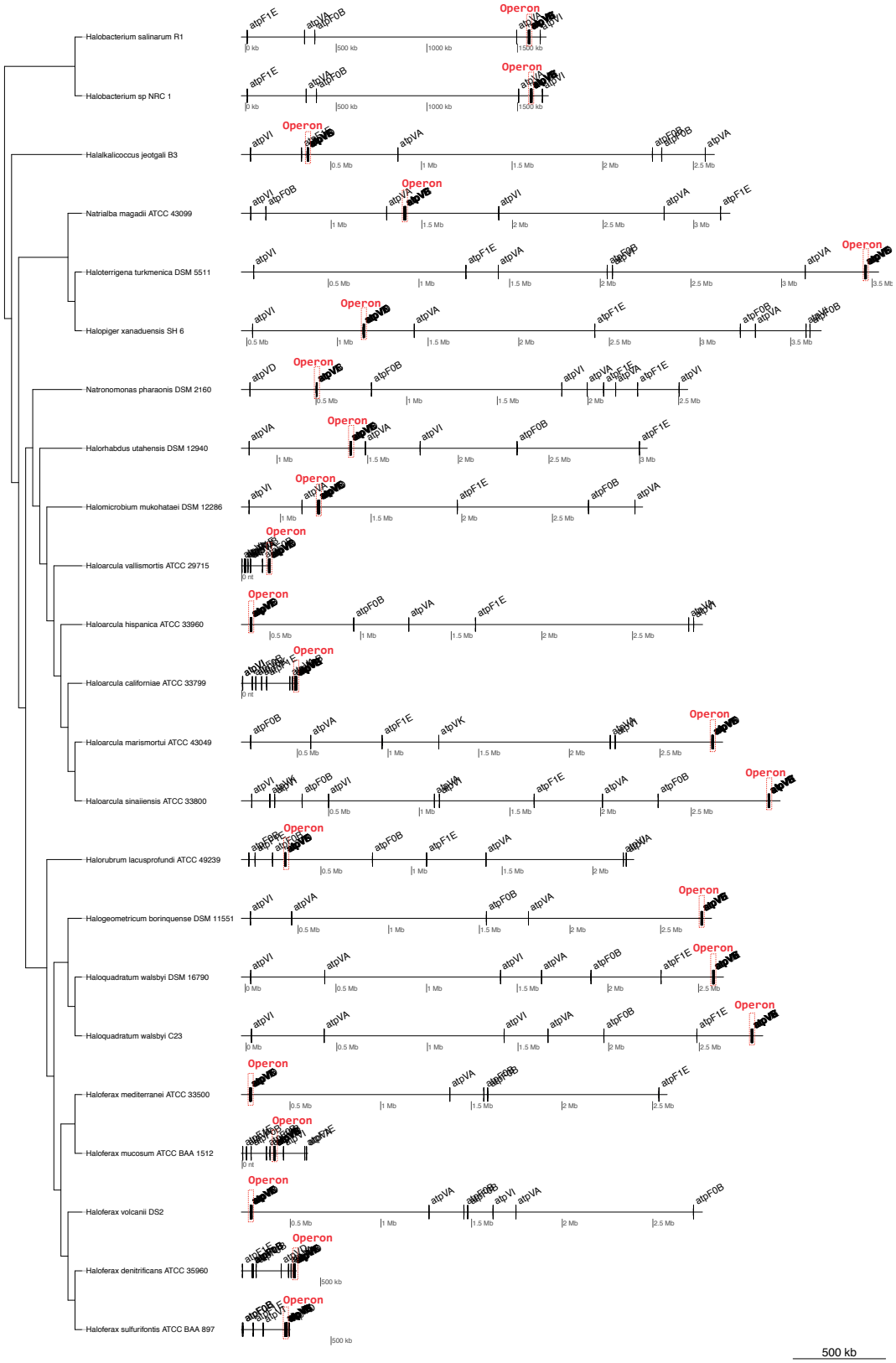


Figure SF17. Distribution genes of the respiratory system with Halobacteriales genomes: Complex V. Genes clustered as operon-like (legends and frames in red), and no clustered (without legends in red) are shown. Halobacteriales genomes are sorted by a subtree obtained from the reference tree reconstructed with concatenated universal proteins for all archaeal genomes (see Chapter 6). Genetic distances are shown in kilobases (kb).

APENDIX II

PHYLOGENETIC TREES

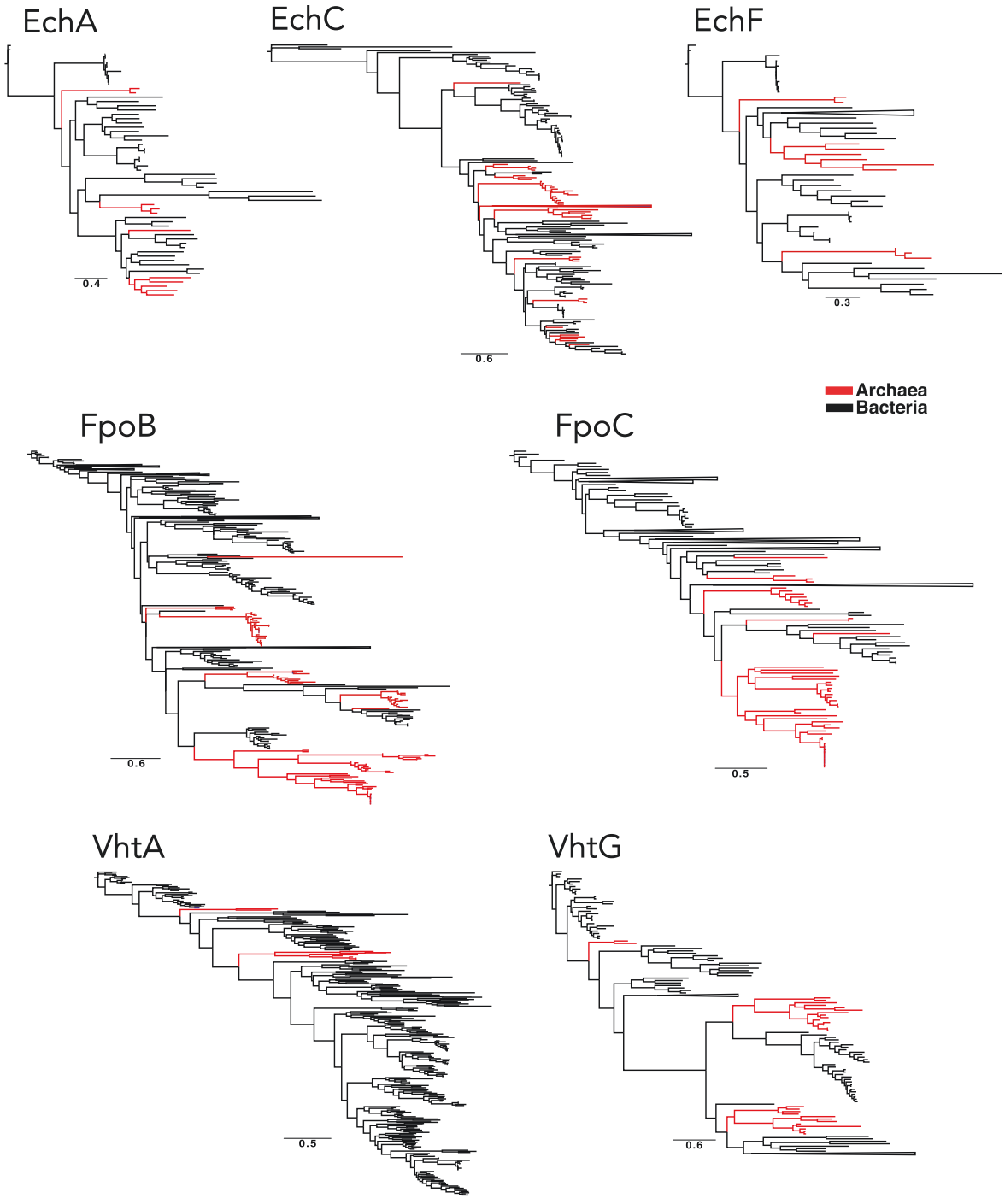


Figure SF18. Undefined HGT events in some of the subunits from the protein complexes of Fpo, Ech, and Vht, as classified in Tale 7.3. Examples of phylogenetic trees from the subunits correspoding protein complexes where the HGT events are classified as undefined transfers, although they seem to be bacterial acquisitions. As observed oneach tree, bacterial clades (branches in black) dominate the tree topology in comparison to the non-monophyletic archaeal clades (branches in red). The density distribution of bacterial species and clades strongly suggest that the HGT has occurred from bacteria to archaea. These trees represent only some subuntis of the complete version of each proteins complex (Ech, Fpo, Vht), in which most of them present (see Table 7.3). Branch distance is shown below each tree.

Acknowledgements

"Mission accomplished, Irma!!
Thanks for all your continuous support, my dear sister."

I would like to express my sincere and deep gratitude to my advisor Prof. Dr. William F. Martin (Bill) for give me the opportunity to develop my doctoral studies in his lab, as well as for his continuous support, encouragement, advice and kind patience during my studies. Likewise, I would like to express my profound gratitude to Dr. Irma Lozada Chávez for her continuous academic and personal support during my studies with insightful discussions and encouragement, even through the most difficult moments of our lives. I'm deeply thankful with Bill and Irma for showing me their passion for science and for their students, as well as for sharing their experience with me has made me a better scientist and person.

In this particular group of people, also I would like to thank to Dr. Filipa L. Sousa for her continuous support, advice and time invested for this thesis, as well as sharing with me her academic experience and valuable friendship.

My sincere gratitude also goes to Prof. Dr. Martin Lercher for being part of the doctoral committee as co-supervisor, and especially to Annette Eder-Martin for her kind support and time invested in all of us (the students of the E-Norm graduate school). Also, I would like to express my gratitude to Doris Matthée and Ariane Baab (retired) for their essential work in the Institute of Molecular Evolution (MOLEVOL).

I couldn't have completed this thesis without the kind support of friends that have helped me at some point in crucial moments during these years. They are: Jens Steuck, Jörn and Tina Habicht, Jörg and Andrea Fallman, Lina Skrickyte, Mario Fergée, Carina Emmel, Cristina Bojórquez Espinosa, Stephanie Kehr, Reine (Byun) Hensgens, Adarelys Andrades and Marco, Tien-Hui Chu (Tien-Tien!!), Gabriel Bauer, Antony Sravan Kumar Reddy, Nina Rossie, Arian Koehler, and Juan Diego (*JD*) Santillana Ortiz.

I thank my fellow labmates in for the valuable moments at MOLEVOL: Dr. Verena Zimorski, Dr. Harald Preisner, Dr. Mayo Röttger, Dr. Sven Gould, Natalia Mrnjavac, Dr. Shijulal Nelson-Sathi, Dr. Thorsten Thiergart, Dr. Christian Wöhle, Peter Melzer, Re-Young Yu, Dr. Gary Kusdian, Sriram Garg, Jan de Vries, Cessa Rauch, Madeline Weis, and Sinje Neukirchen.

Last but not the least, I would like to thank my mother, Ana Maria Chávez Morales, for her support, hope, and unconditional love throughout my entire life, and to my father, Nabor Felix Lozada Guerrero, for his advice and support when a was a child.

This research was supported by the PhD program "Evolutionary Networks: Organisms, Reactions, Molecules (E-Norm)" and by the PhD scholarship for doctorate researcher "STIBET-Studienabschlussbeihilfe" from the Heinrich-Heine Universität Düsseldorf.

"*Nothing last forever… not even our troubles.*"
Arnold H. Glasow

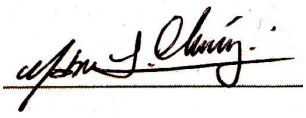
Statement of authorship & Statutory declaration

I hereby certify that this dissertation is the result of my own work. To the best of my knowledge, no other person's work has been used without due acknowledgement.

I declare under oath that I have compiled my dissertation independently and without any undue assistance by third parties under consideration of the 'Principles for the Safeguarding of Good Scientific Practice at Heinrich Heine University Düsseldorf'.

This dissertation has not been submitted in the same or similar form to other institutions. I have not previously failed a doctoral examination procedure.

Düsseldorf, 14.05.2018



Alejandro Nabor Lozada Chávez