Une histoire de deux membranes. Transition de l'enveloppe cellulaire des Firmicutes

Daniel Poppleton

Université Pierre et Marie Curie

Complexité du Vivant

Biologie Moléculaire du Gène Chez les Extrêmophiles, Instiut Pasteur

A Tale of Two Membranes

Cell Envelope Transition in Firmicutes

Daniel Ian Poppleton

Thèse de doctorat de biologie

Dirigée par Simonetta Gribaldo

Présentée et soutenue publiquement le 12 Octobre 2017

Devant un jury composé de :

Guennadi Sezonov	Professeur à l'UPMC	Président du Jury
Tâm Mignot	Directeur de Recherche	Rapporteur
Frederique Le Roux	Directeur de Recherche	Rapporteur
Felipe Cava	Professeur associé	Examinateur
Waldemar Vollmer	Professeur	Examinateur
Maria Lina Bernardini	Professeur	Examinateur
Christophe Beloin	Directeur de Recherche	Membre invité
Simonetta Gribaldo	Directeur de Recherche	Directeur de these

For my wife

ACKNOWLEDGEMENTS

There are many people who made this work possible. The first and most important is my wife, Christina. Her unending support and brilliance enabled me to come and has been as essential throughout my whole PhD. My supervisor, Simonetta, has been ideal through this PhD. She has helped even when I have been difficult and offered me key advice. My unofficial secondary supervisor Christophe has been just as superb with a more relaxed approach and both deserve strong thanks. Our lab group has one of the best secretaries in history. I can not overstate how much she has helped a dumb Anglophone cut through and understand French administration. The unit of BMGE has excellent members, all of whom have contributed to my PhD, however, there are a few that helped me immensely. The first is Patrick as he has taught me new ways to look at life and evolution and the final chapter is a tribute to his ideas. Although, he will probably disagree with most of it…. Our laboratory group has been excellent with strong communication and synergy. Within it, Luisa, Guillaume, and Panos have helped me produce amazing work. My close friends Morgan and Elena have been helpful in small details and broader perspective. Morgan has taught me that the most difficult path may not be the most exciting and fulfilling, while Elena has trained me on the culture of France with some fuming discussions. Finally, I would like to thank my family. Mom and Dad for believing in me, even though they understand little about the subject and care about it even less. Kirsty and Christopher for being there when I need them. Thanks, Jon and Jan for showing up and having great interest. There are many others who have enabled me to produce this work and I apologize if I forgot you.

Thank you all

TABLE OF CONTENTS

TABLE OF ABBREVIATIONS

Abbreviations used throughout the text. Common chemical abbreviations will not be introduced in the text.

Abbreviation	Expansion
ATP	Adenosine Triphosphate
BAM	ß-barrel Assembly Apparatus
CM	Cytoplasmic Membrane
CPR	Candidate Phyla Radiation
DCW	Division and Cell Wall
DNA	Deoxyribose Nucleic Acid
Glc	Glucose
HMM	Hidden Markov Model
HMW	High Molecular Weight
IM	Inner Membrane (synonymous with CM)
KDO	keto-Doxyoctulosonate
LCA	Last Common Ancestor
LPS	Lipopolysaccharide
LT	Lytic Transglycosylase
LTA	Lipoteichoic Acids
LUCA	Last Universal Common Ancestor
MGT	Monofunctional Glycosyltransferases
ML	Maximum Likelihood
NAG	N-acetylglucosamine

Abbreviation	Expansion
NAM	N-acetylmuramic acid
OM	Outer Membrane
OMV	Outer Membrane Vesicle
PBP	Penicillin Binding Protein
PEP	Phospho-Enol-Pyruvate
PG	Peptidoglycan
PM	Plasma Membrane (synonymous with CM)
RNA	Ribose Nucleic Acid
SAM	Sorting and Assembly Machinery
T2SS	Type 2 Secretion System
T3SS	Type 3 Secretion System
T4P	Type 4 Pili
TAM	Translocation and Assembly Module
TPS	Two Partner Secretion System
TVSS	Type V Secretion System (synonymous with autotransporter)
UDP	Uradine Diphosphate
WTA	Wall Teichoic Acid
xxxase	cleaves xxx bond

FOREWORD

*It was the best of envelopes, it was the worst of envelopes. It was the age of adaptation, it was the age of extinction. It was the epoch of evolution, it was the epoch of stasis. It was the season of Protein, it was the season of DNA. It was the spring of secretion, it was the winter of ingestion. We had everything protecting us, we had nothing protecting us. We were all going directly to mesophillia, we were all going direct the other way. In short, the period was so far like the present period, that creatures great and small were evolving, and we had no idea what was happening.**

The discovery and observation of single cell organisms were performed by Antoni van Leeuwenhoek in the mid-17th century (Leewenhoeck, 1677). He observed the presence of "small animals" in his spittle and the spittle of other subjects. He observed (Figure F.1) what we now assume to be bacilli (Figure F.1.A), micrococci (Figure F.1.E), and *Selenomonas* (Figure F.1.B). The *Selenomonas* are of great interest to this manuscript as they are Negativicutes, a relatively unknown clade discussed in the introduction section 4.1, which has been at the forefront of great bacterial discovery.

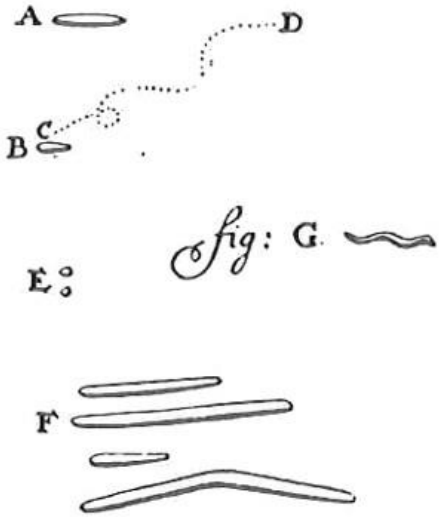

Figure F.1. First Observation of Bacteria by Antoni Van Leeuwenhoek.
Taken from (Leewenhoeck, 1677).

The next major step forward in the study of bacterial cell envelopes occurred in the late 19th century when Hans Christian Gram discovered a mechanism for differentiating

**Modified from "A Tale of Two Cities" by Charles Dickens*

bacteria from the lungs of pneumonia victims (Gram, 1884). This simple four step method provided easy visual differentiation of bacteria. Whereby, if the bacteria absorbed the stain they were identified as Gram positive (*Bacillus subtilis* for example) and if they do not they are Gram negative (*i.e*. *Escherichia coli*).

It took 50 years more until it was possible to properly determine why these two groups of bacteria are different. In a hallmark paper by (Bladen and Mergenhagen, 1964) they succeeded in capturing the first image of a double membraned bacteria. It was not *E. coli* as you would expect, but another Negativicute, *Veillonella parvula* (Figure F.2). Since then work has been done on describing the protein systems, membrane composition, and basic physiology of both double and single membrane bacteria. However, no one has yet definitively demonstrated the ancestry of one system or another. This dissertation will focus on the ancestry of cell envelopes and provide evidence for one system's ancient origins.

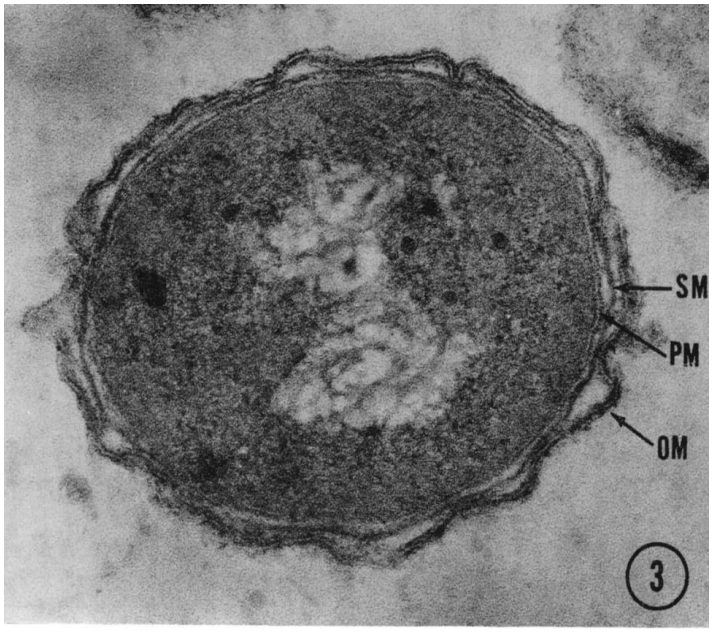

Figure F.2. First Observation of a Bacterial Outer Membrane.
Image of *V. parvula* showing the OM, peptidoglycan (SM or structural membrane), and cytoplasmic membrane (PM or plasma membrane). Taken with permission from (Bladen and Mergenhagen, 1964).

INTRODUCTION

1 Diderm and Monoderm Cell Envelope Architecture

The basic function of bacterial cell envelopes is to separate the intracellular cytoplasm from the extracellular medium. In addition, prokaryotes utilize this structure to provide a framework of the cell. Most components of these envelopes are necessary for providing this structural barrier or circumventing it. The structure provided by peptidoglycan (PG) or the transport mechanisms for solute entry are two such examples. Bacteria can be divided into two groups depending on whether they possess one membrane, as in monoderm, or two, as in diderm. The mechanisms of transport and structure are often different between diderm and monoderm bacteria and as such, to understand the origin and division of these cell envelopes we must first discuss the basic composition and systems of both monoderm, exemplified here by *B. subtilis*, and diderm envelopes, exemplified here by *E. coli* (Figure I.1).

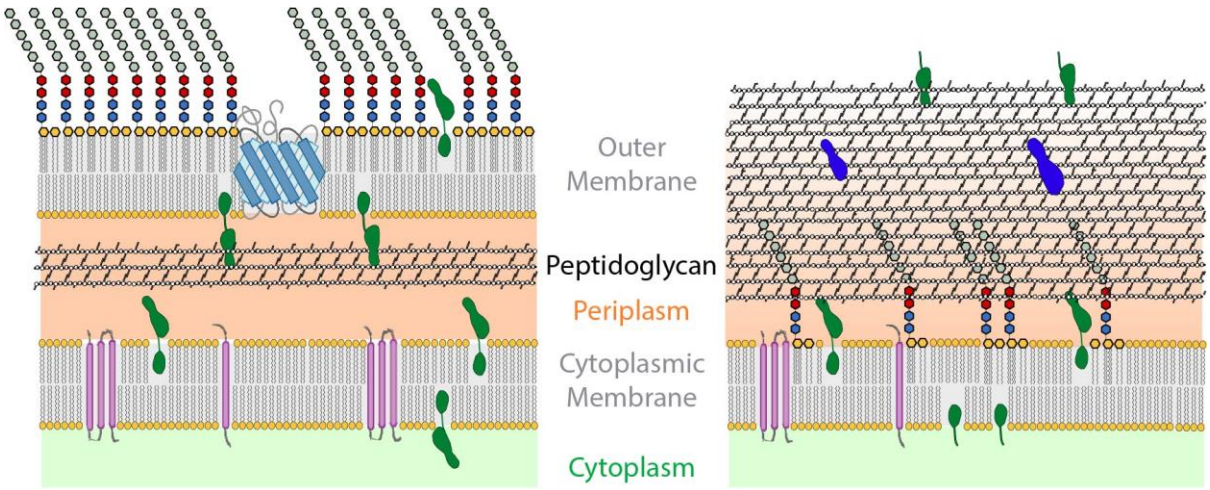

Figure I.1. Diderm vs Monoderm Cell Envelope Architecture.
Both Diderm (Left) and monoderm (right) share a cytoplasmic space and cytoplasmic membrane, however, they differ in that diderm bacteria possess an outer membrane with lipopolysaccharides within the outer leaflet. In addition, diderm bacteria possess a thin layer of peptidoglycan while monoderms have a thick layer.

1.1 Diderm vs. Monoderm Broad Characteristics

Monoderm and diderm bacteria have far more similarities than differences; In general, they have the same machinery for protein synthesis, DNA replication, lipid biosynthesis, and basic metabolism. The key differences lie past the cytoplasmic membrane (Figure I.1). Monoderm bacteria are so named due to possessing only a single cytoplasmic (CM), while diderm bacteria have an outer membrane (OM) not found in the other domains of life. The CM of monoderm and diderm bacteria are homologous, however, they are frequently called the inner membrane (IM). For the sake of this document the more accurate term of CM will be used whenever possible, however, you may see the synonymous PM (Plasma Membrane) or IM. The predominant similarity between diderm and monoderm cell envelopes is the CM. Distal from this membrane both bacteria have PG (discussed in-depth in introduction section 1.2), however, it is different between the two envelope types. The most striking difference in the PG is that in diderm bacteria the PG forms a thin exoskeleton (1-4 molecules thick) located in the periplasm, the space between the CM and the OM, while monoderm bacteria usually have thick PG which is 30-40 times the thickness of diderms PG (Vollmer *et al.*, 2008). There are many diderm exclusive systems and machinery (See next section), while conserved monoderm exclusive systems are rare (Introduction Section 1.4).

1.2 Peptidoglycan

An essential cell envelope component, the PG, has been one of the defining differences between monoderm and diderm bacteria, with diderm having a thin PG layer and monoderm possessing a thick PG layer. This difference in width is a historical distinction between diderm and monoderm bacteria, studied with the model organisms *E. coli* and *B. subtilis* respectively. In *E. coli* the PG is thin, 2-7 nm, while in *B. subtilis* and other monoderm bacteria it is thicker and ranges from 20-80 nm. PG is found in almost all bacteria (Schleifer and Kandler, 1972) and is involved in several key functions such as cell shape, osmotic pressure (Turner *et al.*, 2014), size (Schoonmaker *et al.*, 2014), division (Jacquier, Viollier, & Greub, 2015), and antibiotic susceptibility and resistance (Jacquier *et al.*, 2015). Aside from these key functions, the PG sacculus has been implicated in most physiological processes of the bacterial cell (Cava and de Pedro, 2014). In both diderm and monoderm bacteria, PG consists of a mesh-like polymer constituted of repeating disaccharides alternating between N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM) (Figure I.2). This glycan chain is crosslinked through a 3-5-residue peptide chain attached to the NAM. The basic structure is conserved throughout bacteria however, monoderm bacteria have great variability in both the peptide chain and the type of cross-link (Schleifer and Kandler, 1972).

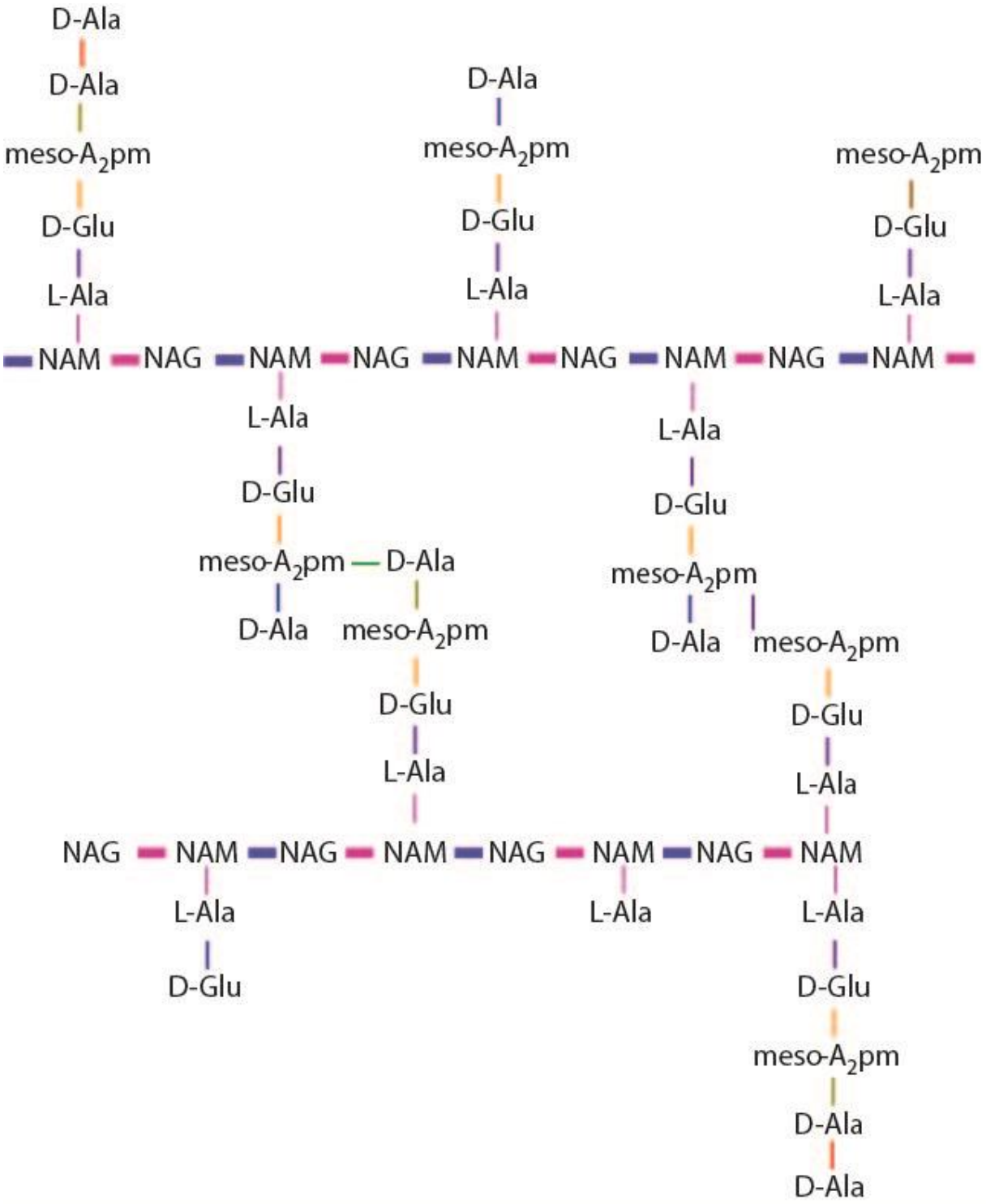

Figure I.2. The Structure of Peptidoglycan.
Peptidoglycan includes a sugar backbone (NAM & NAG), a peptide side-chain, and crosslinks between the side-chains.

1.2.1 Biosynthesis

PG biosynthesis is accomplished through a conserved pathway found in nearly all bacteria. This biosynthesis can be separated into three steps based on the localization of the enzymes (Figure I.3). To begin, a precursor is anabolized in the cytoplasm, and then it is lipidated and flipped through the CM. After the PG monomer is flipped it is polymerized into a chain and attached to the existing PG sacculus.

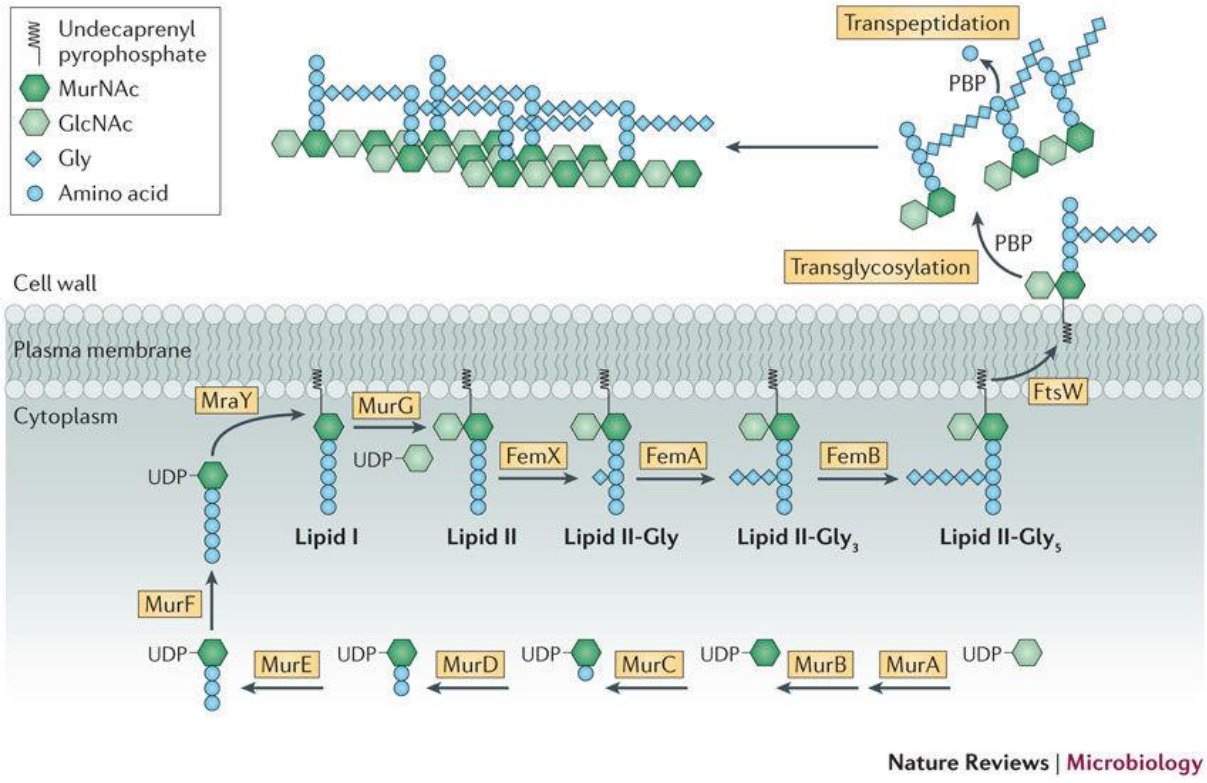

Figure I.3. Peptidoglycan Biosynthetic Pathway.
Peptidoglycan is synthesized in a series of cytoplasmic steps using MurA-F to build the basic precursor. After a series of steps occur on the inner leaflet of the CM, it is then flipped via a translocase and added to the sacculus with penicillin binding proteins that catalyze the transpeptidation and transglycosylation reactions. Taken with permission from (Pinho *et al.*, 2013)

The initial precursors are synthesized from Fructose-6-phosphate (F6P) to produce UDP-Glc-NAG using the glucosamine pathway. UDP-Glc-NAG is utilized for the production of chitin, glycoproteins, and PG. The first committed step in PG biosynthesis is the addition of a PEP to UDP-Glc-NAG by MurA. The molecule is then reduced by MurB to prepare the

molecule for amino acid addition and hydrolysis. Then the structurally homologous enzymes MurC, MurD, MurE, and MurF sequentially add amino acids. The substrates for each of these enzymes are a single amino acid, except for that of MurF, which adds a dipeptide produced by the dimerization enzyme Ddl. Each of these four enzymes has a conventional amino acid that is added however, the amino acids can be different between clades of bacteria or in an individual strain to produce antibiotic resistance (Barreteau *et al.*, 2008).

After MurF produces the PG subunit a series of lipidation reactions occurs at the CM. MurG is responsible for lipidating the precursor producing lipid I. MraY then adds a second sugar, GlcNAC, to form lipid II (Bouhss *et al.*, 2008). The final step before addition to the PG is flipping Lipid II to the outer leaflet. There is a debate on which enzyme is responsible for this process. Two candidates have been put forward, MurJ and FtsW; strong evidence supports each protein, however, these studies have conflicting results (Mohammadi *et al.*, 2011; Sham *et al.*, 2014).

After transport through the CM, lipid II is polymerized and cross-linked to the mature PG. This process is performed by high molecular weight (HMW) penicillin-binding proteins (PBP). Polymerization is performed by Class A HMW PBPs. Class A PBPs consists of two domains: a membrane-linked transglycosylase domain and a transpeptidase domain. The transglycosylase domain sequentially adds a Lipid II to a chain producing Lipid IV, Lipid VI, Lipid VIII, etc. The transpeptidase domain creates crosslinks with the existing PG structure. The chain is then further crosslinked by Class B HMW PBPs, which do not have a transglycosylase domain, rather they have a dimerization domain that is not well characterized. The chain length is variable in bacteria and the process of chain termination is unknown (Sauvage *et al.*, 2008a).

1.2.2 Genetic Organization

The genetic localization of most PG synthesis genes is located in a conserved region; the Division and Cell Wall (DCW) cluster (Figure I.4). This cluster contains three to four functional sub regions. The first subregion contains *mraZ*, *mraW*, and *ftsL*. MraZ and MraW are responsible for negative regulation of cell division including the DCW cluster (Eraso *et al.*, 2014). FtsL connects the division complex of the cytoplasm with the division complex of the periplasm. Additionally, *ftsL* is upstream of the PG biosynthesis genes *murA-G*, *ddl*, *ftsW*, and *mraY*. After the PG biosynthetic sub region a second division sub region is found,

containing *ftsQ*, *ftsA*, and *ftsZ*; all essential genes for division. This cluster is well conserved in all PG producing bacteria with typical division (Eraso *et al.*, 2014).

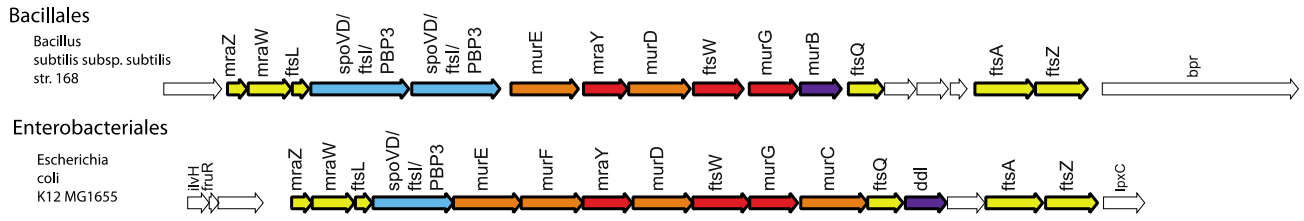

Figure I.4. The Division and Cell Wall Cluster
Genetic organizations of the DCW cluster in a Firmicute, *B. subtilis*, and a Proteobacteria, *E. coli*.

1.2.3 Remodeling

To grow, divide, and insert protein complexes in the periplasm the PG must be remodelled. This is a complex process involving multiple classes of enzymes that create and cleave the bonds of the PG named autolysins or PG hydrolases. These enzymes are difficult to classify, as there is a high number present in most genomes and many have redundant or multiple roles. Nonetheless a biochemical classification does exist based on the types of bonds that are cleaved (Figure I.5).

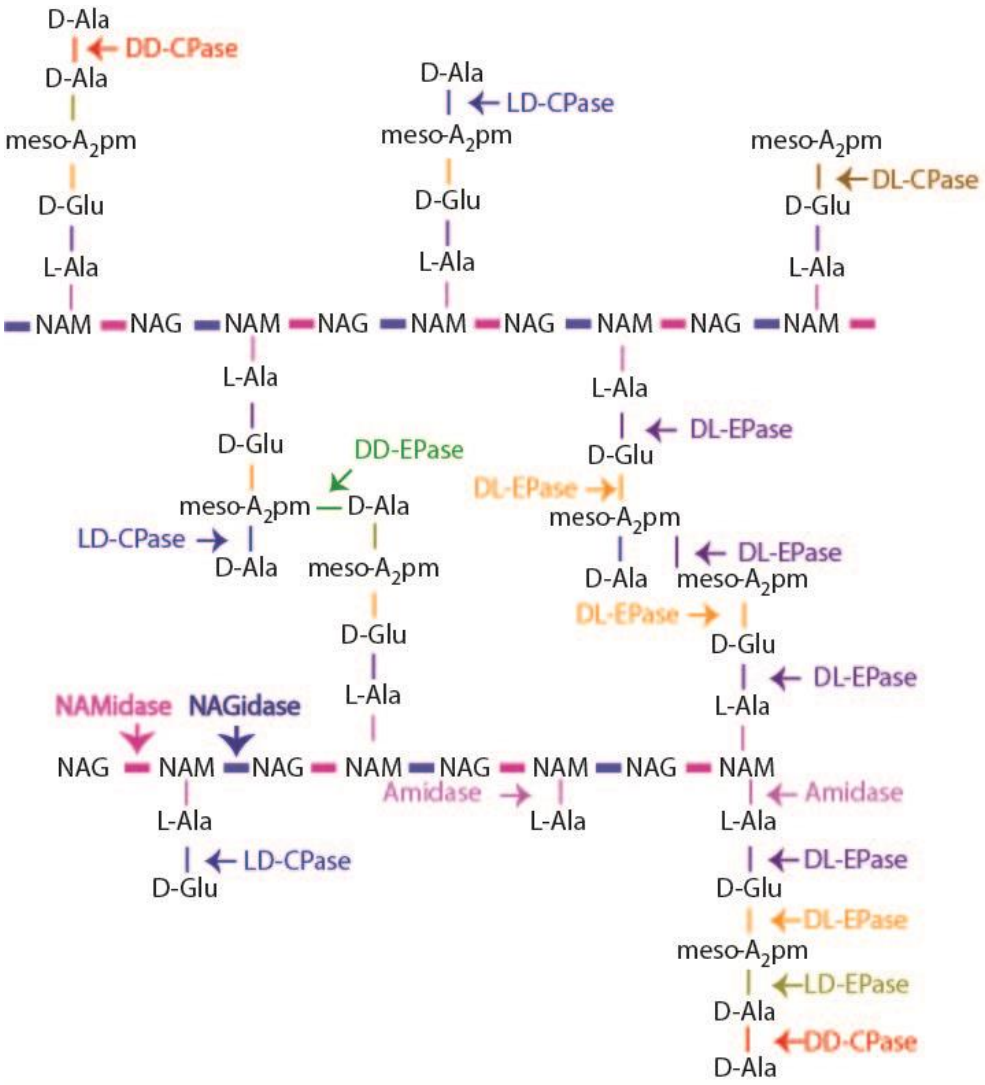

Figure I.5. Peptidoglycan Hydrolases and Cleavage Sites.
Peptidoglycan hydrolases are defined on where the type of bond they can cleave, each different bond is shown with a different colour corresponding to the enzyme type that cleaves the bond.

Four different types of enzymes can cleave the sugar backbone of PG; lysozyme, lytic transglycosylases (LT), N-acetylglucosaminidase (NAGidase) and the N-acetylmuramidase (NAMidase). Lysozymes are a non-specific hydrolase that can cleave anywhere on the PG background and is found in many species. It has been classified into Chicken-type, Goose-type, phage-type, bacterial-type, and fungal-type. Each group represents a distinct phylogenetic clade and all lysozymes share structural homology. Most bacteria lack lysozyme as it is not used for controlled PG remodelling, those that do produce it use it in bacterial warfare (Pei and Grishin, 2005). Lytic transglycosylases cleave PG at the same location, the β-1, 4-glycosidic bond between NAM and NAG, yet produce an anhydromuramoyl product rather than the hydrolyzed sugar. This has an effect of terminating the PG chain at this point and preventing further additions. The final type is the NAGidase and NAMidase, which

cleave in a mechanism similar to that of lysozyme except they have much more specific activities; NAGidases hydrolyze the glycosidic bond between an N-acetyl-β-D-glucosamine residue and the adjacent monosaccharide.

N-Acetylmuramyl-L-Ala amidases (NAM-LAAs) are an essential enzyme in remodelling, which split the bond between the peptide side chain and the backbone of PG. These enzymes are used extensively by phages for lysis of the host cell and lack signal peptides to cross the CM. The peptide side-chain of the PG undergoes extensive remodelling during growth, division, and sporulation. This remodelling is accomplished through the use of endopeptidases (EPase) and carboxypeptidases (CPase). EPases cleave in the centre of a peptide chain or at the crosslink attaching two chains. CPases remove the terminal residue from a string of amino acids.

1.3 Diderm Characteristics and Systems

As diderm bacteria possess a relatively impermeable barrier to most molecules, complex machinery must be present to transport molecules in both directions past the membrane. Furthermore, proteomic machinery is necessary to assemble the membrane and its machinery. The key components of the outer membrane relevant to this document will now be discussed in detail.

1.3.1 Lipopolysaccharides

Lipopolysaccharides (LPS) or endotoxins are the major component of the OM in diderm bacteria, comprising the entire lipid composition of the outer leaflet of the OM in most diderm clades. LPS is incredibly important in human health; not only is it a toxin responsible for endotoxin shock, it is also the ligand of Toll-like receptor 4 on immune cells (Sveen *et al.*, 1977; Matera *et al.*, 2009). On the opposing side, the LPS of Cyanobacteria has been shown to have a positive impact on human health (Durai *et al.*, 2015).

1.3.1.1 LPS Structure

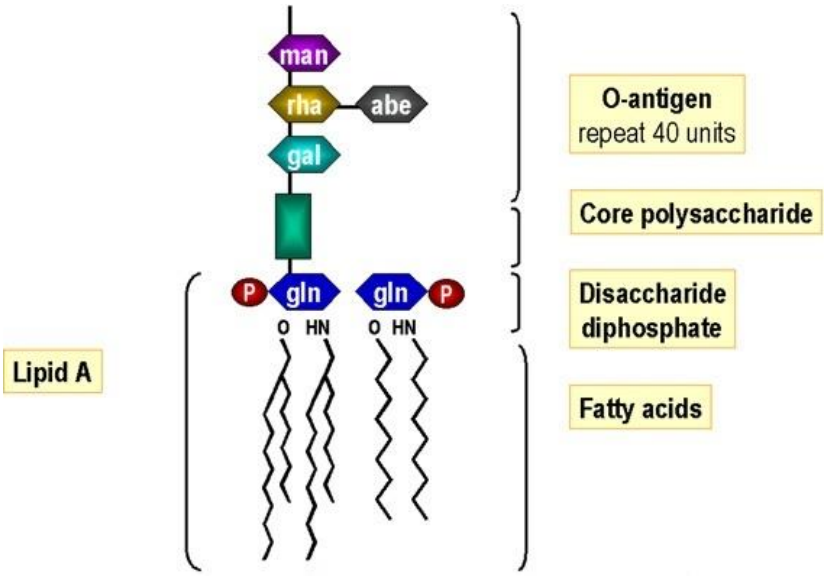

Figure I.6. The structure of LPS.
LPS contains a hydrophobic Lipid-A moiety consisting of fatty acids and a disaccharide diphosphate, then extending externally is the core polysaccharide, and finally the polymeric O-antigen.

Each individual LPS molecule consists of three functional regions: O-Antigen, Core oligosaccharide, and Lipid A (Figure I.6). LPS is described as smooth if it contains all three components, rough if it only contains Lipid A as well as the core oligosaccharide and deep rough if only Lipid A is present. Lipid A is the most fundamental component and is essential in many diderm bacteria (Polissi and Sperandeo, 2014). It consists of a phosphorylated glucosamine disaccharide which has multiple fatty acids covalently attached. The fatty acids constitute the hydrophobic portion of the outer leaflet of the OM, while the phosphor-disaccharide constitutes the hydrophobic portion. This provides a functional analogue to phospholipids however, Lipid A is much larger. Although this moiety can exist without any further modification, as is found in many cyanobacteria (Opiyo *et al.*, 2010), normally additional molecules are added.

The further the distance from the hydrophobic fatty acid portion, increased diversity can be found. In most bacterial species Lipid A is covalently attached to a keto-doxyoctulosonate (KDO) producing the core oligosaccharide, however, this KDO is usually modified with additional saccharides and can be further modified with phosphate groups, amino acids, sugars, polyamines, or acidic molecules (Needham and Trent, 2013). One of the most diverse antigens found in nature is the final component of LPS, O-antigen. In fact, O-antigen has been used to type bacterial strains since the 1940's with over 196 different types currently identified for *E. coli* alone (DebRoy *et al.*, 2016). As an example, *E.coli* O157: H7 belongs to the 157 antigen group of LPS. This incredible diversity is produced through duplication of small modifications. It can consist of many different monomeric subunits, however, all O-antigens are polymers to some extent. The length of this polymer varies from strain to strain.

1.3.1.2 KDO-LPS Biosynthesis

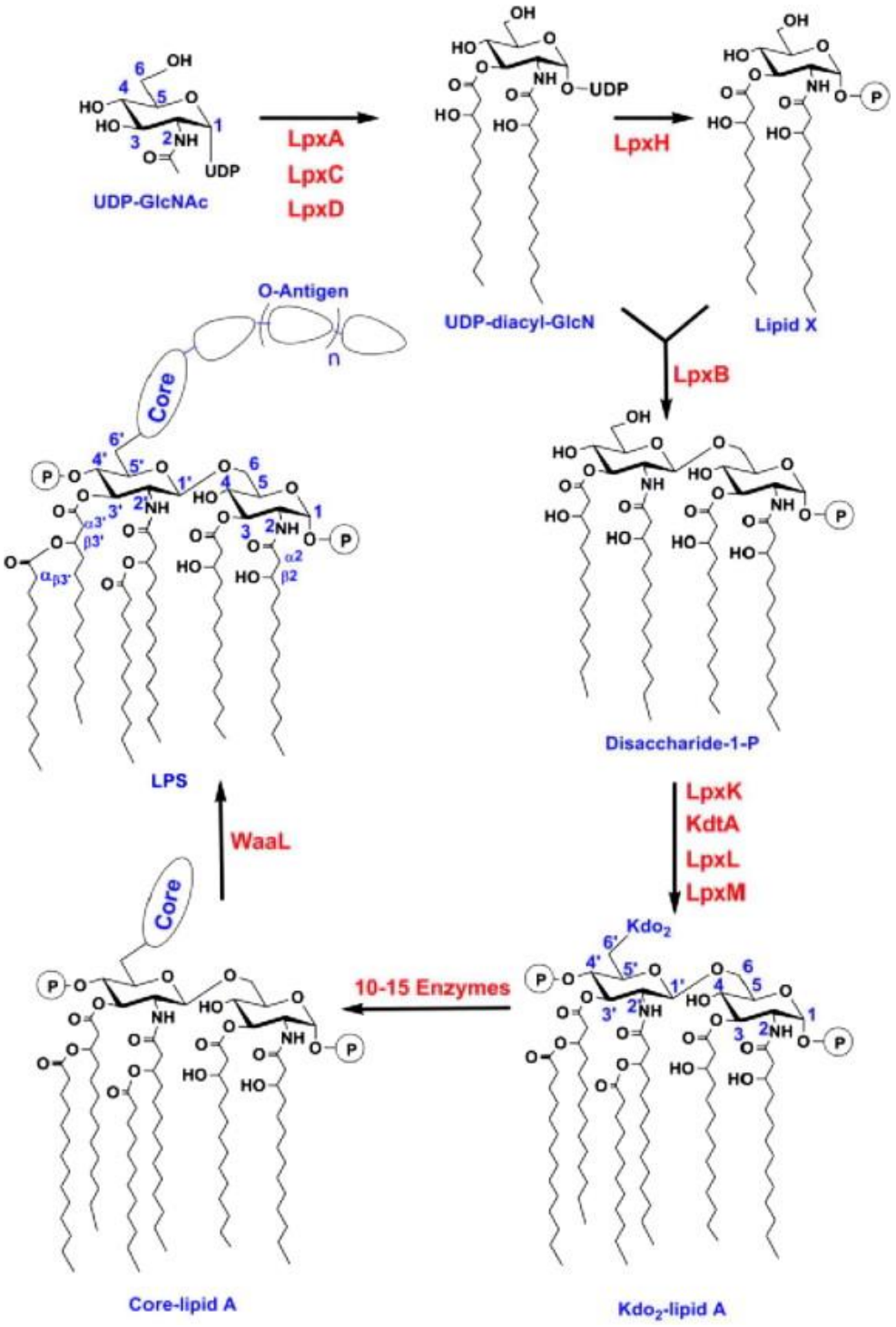

Figure I.7. The Biosynthetic Pathway of LPS.
The biosynthetic pathway of LPS starts with the attachment of fatty acids to a UDP-GlcNAc sugar, then a molecule is phosphorylated and the two molecules are connected. After this a KDO molecule is added, then the core polysaccharides are added and finally, O-antigen is attached. Taken with permission from (Wang and Quinn, 2010)

Of special importance to the data presented in this thesis is the biosynthetic pathway of LPS (Figure I.7). First, a UDP-GlcNAc is lipidated by the three soluble enzymes LpxA, LpxC and LpxD. This produces a lipidated sugar, which is phosphorylated by LpxH and then dimerized by LpxB. This dimerization produces the Lipid A molecule. Some bacteria stop here and transport the molecule in this state. Most bacteria, however, produce KDO-Lipid A, which requires the integral CM proteins LpxK, L, M, and KdtA. LpxK first phosphorylates the LPS moiety, after which KdtA attaches two KDO residues to the molecule, producing KDO-Lipid A. Following this two fatty acids are attached by LpxL and LpxM producing the base molecule (Wang and Quinn, 2010). Due to the diverse nature of O-antigen and core oligosaccharide structure and corresponding biosynthesis, only the core biosynthetic steps of Lipid-A biosynthesis will be analyzed and discussed in this thesis.

1.3.1.3 LPS Transport

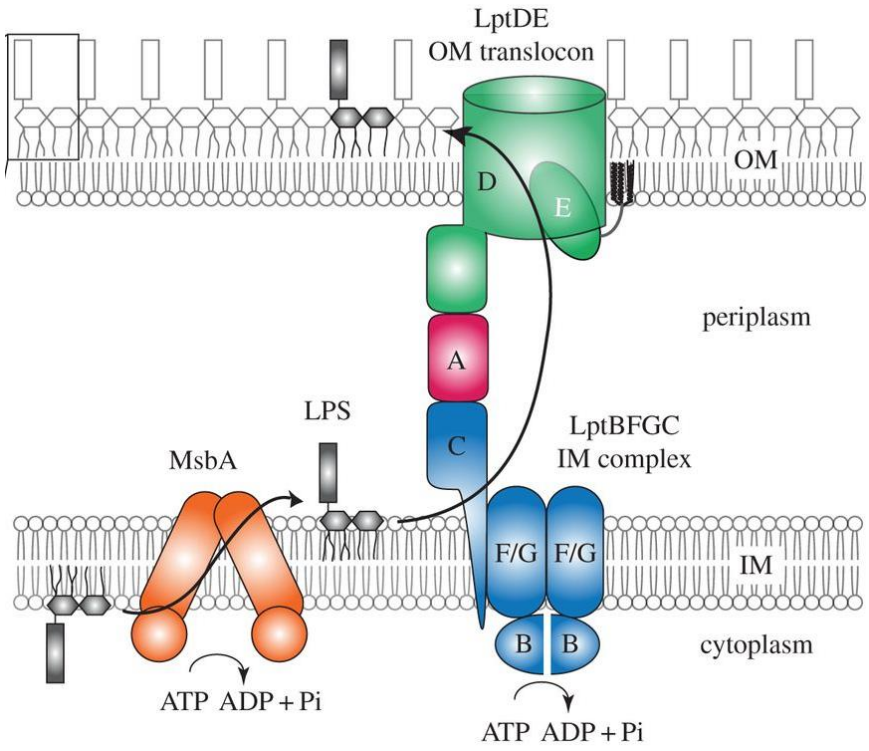

Figure I.8. OM Transport of LPS.
The lipid-A molecule is first flipped to the outer leaflet of the CM and then transported along a chain of carrier proteins (LptA) by the energy from a CM ATPase (LptBFGC) to the OM translocon (LptDE). Taken with permission from (Simpson et al., 2015).

Distinct from O-antigen transport, which has a dedicated pathway, Lipid-A is transported with the Lpt pathway (Figure I.8). Before the Lpt pathway can act, the LPS

monomer is flipped to the outer leaflet of the CM by MsbA. The pathway is energized by a CM ATPase consisting of LptBFGC and the molecule is transferred to the periplasmic carrier LptA. LptA then transfers the molecule to the OM receptor LptD, a ß-barrel, and LptE, the cap protein of LptD. The molecule is then incorporated into the OM (Wang and Quinn, 2010).

1.3.2 Flagella

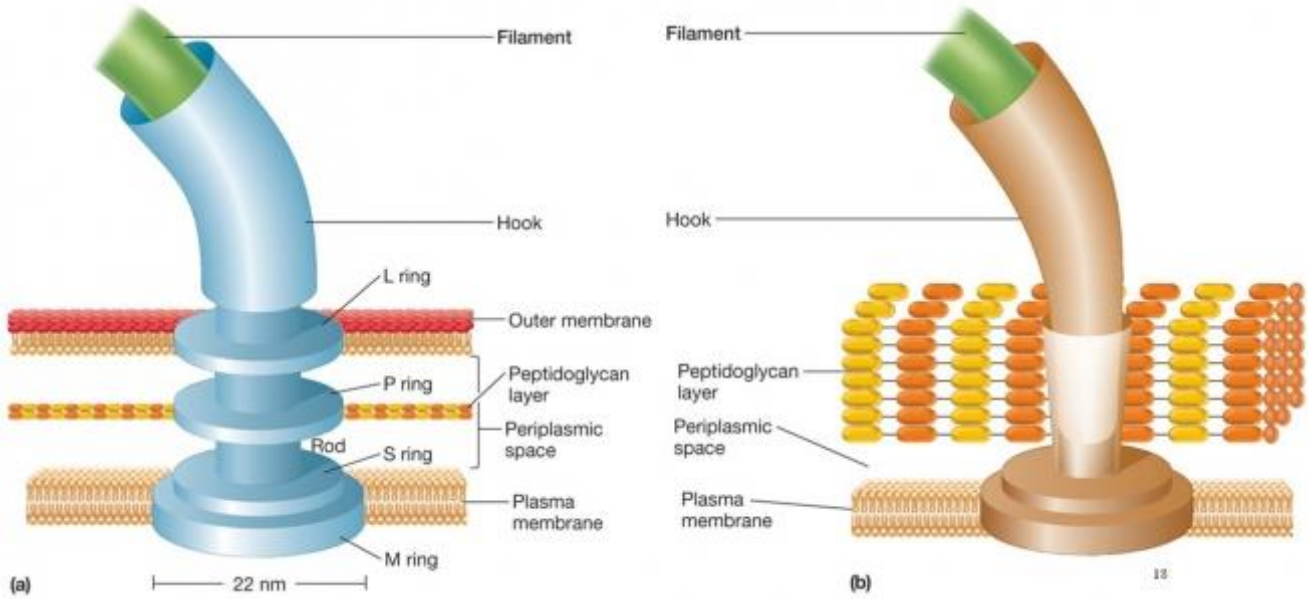

Figure I.9. Diderm and Monoderm Flagella.
The bacterial flagella of diderms (left) and monoderms (right). Notice the primary difference is the absence of the P and L ring in the monoderm type.

The bacterial flagellum (Figure I.9) is a whip-like appendage that extends from the envelope of many bacteria for motility. This structure is found in both diderm and monoderm bacteria. The structures are predominantly similar, however, there are a few key differences. The components shared among diderm and monoderms form a complex structure that is assembled with an entourage of 34+ proteins and as such will not be discussed in detail. For a detailed review please see (Mukherjee and Kearns, 2014). What is important to discuss is the differences between the monoderm type and the diderm type found in Bacteria. In general, the diderm flagella and monoderm flagella are nearly identical and share components amongst themselves with the exception of the diderm flagella having a few extra proteins. These proteins are FlgA, FlgH, and FlgI. FlgI is the P-ring protein, which mechanistically functions as a bushing that separates the spinning flagella from the skeletal PG layer. FlgA is a chaperone for FlgI, aiding in its assembly and preventing proteolytic degradation by the native residents of the periplasm. FlgH also has a bushing function, however, it resides in the

OM/LPS. Aside from the absence of these proteins in monoderm bacteria, genetic clustering of the flagellar genes are more prevalent in the monoderm groups of Actinobacteria and Firmicutes (Mukherjee and Kearns, 2014).

1.3.3 Secretion Systems of Diderm Bacteria

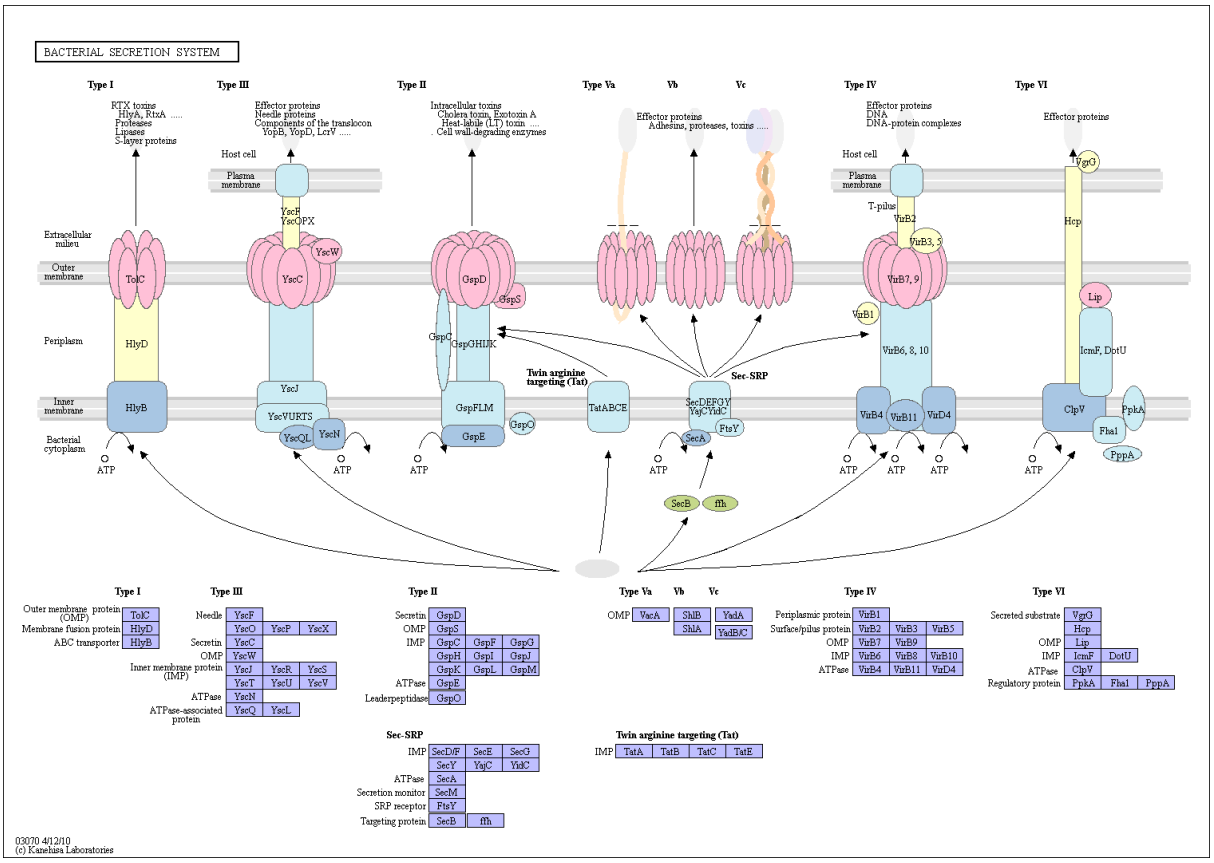

Figure I.10. Well Characterized Secretion Systems of Diderm Bacteria.
Hypothesized structure of all well-known systems that are used for translocation of proteins past the OM (A). Names for all major components of each system (B). Figure obtained from KEGG (Kanehisa *et al.*, 2017)

There are currently nine secretion systems for the dedicated transfer of substrates past the outer membrane in diderm bacteria (Green and Mecsas, 2016), of which six have been extensively studied (Figure I.10). These systems are known as TXSS with the X referring to the number of the system. Many of these systems have unique components that enable transport of proteins past the OM that are not found in monoderm bacteria, however, few studies have looked at these systems in monoderm bacteria. In the following sections, I will discuss the systems of relevance to this thesis, namely Type 1, 2, and 5. For a recent review and full description of the other systems please see (Abby *et al.*, 2016), which also covers the distribution of these systems among Bacteria.

1.3.3.1 Type I Secretion System

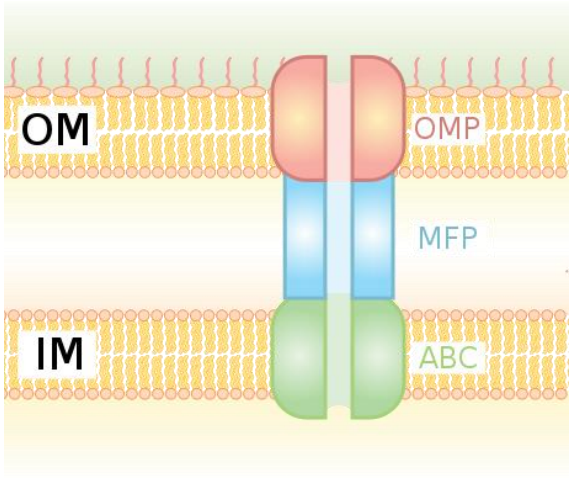

Figure I.11. The Type I Secretion System.
The T1SS consist of a CM ABC transporter (Green), a periplasmic membrane fusion protein (Blue), and an outer membrane channel (Red) forming a single mono-directional ATP driven channel.

Type I secretion system (Figure I.11), or TolC transport, is a SEC independent mechanism of transport. It bypasses the periplasm and secretes from the cytoplasm directly to the extracellular space. T1SS utilizes ATP to drive expulsion of proteins with a size range of 10-250 kDa, toxins, and small substrates. It consists of a CM ABC transporter, a periplasmic channel protein, and an OM channel. For a review please see (Buchanan, 2001).

1.3.3.2 Type IV Pili/ Type II Secretion System

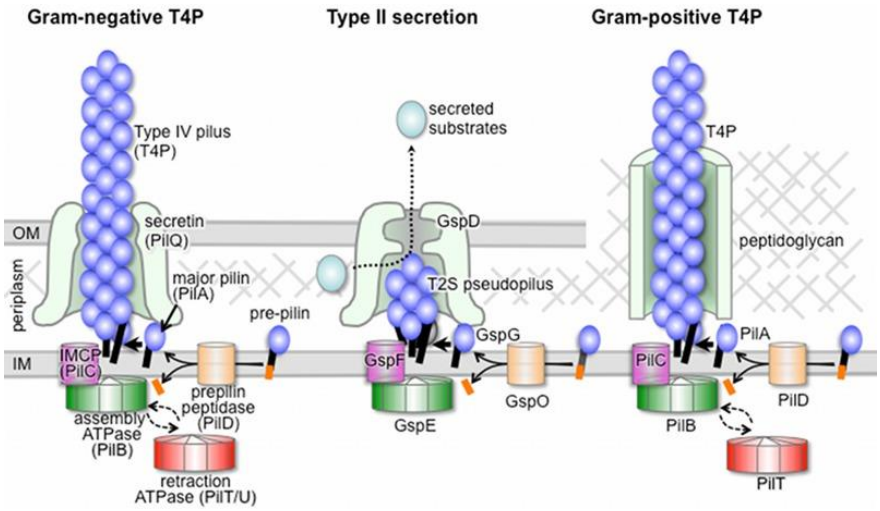

Figure I.12. Type II Secretion System and Type 4 Pili of Diderms and Monoderms.
The T2SS/T4P consist of a CM platform for hydrolyzing ATP to transport substrate/extend pilins and an OM pore found only in diderm bacteria.

The type 2 secretion system (T2SS) (Figure I.12) is one of the major pathways utilized by diderm bacteria in the transport of folded enzymes and toxins from the periplasm to the extracellular space. The T2SS complex is composed of 12-16 proteins, which extend to the extracellular environment. The proteins exported by this pathway are fully folded before leaving the cell. These substrates include toxins such as cholera toxin and hydrolytic enzymes including lipases and proteases. For a recent review see (Zückert, 2014). This system is homologous to the type 4 pili (T4P) that many bacteria use for adhesion, cohesion, twitching motility, and competence in both diderm and monoderm bacteria.

The T2SS and T4P share many of the same protein systems, which are used for function. The main components are the pilin subunits, PilA, which are cleaved using a pre-pilin peptidase, PilD. The pilin subunits are bundled on a CM platform, PilC, and extend past the OM through a pore protein or secretin, PilQ, not found in most monoderm bacteria. The pilus has the unique feature of extension and retraction by the ATPase's PilB and PilT respectively. T2SS lack the retraction ATPase and their pili never penetrate the secretin; they function in the periplasmic space.

1.3.3.3 Secretin

To keep equilibrium within a cell, transport must be precisely controlled. Most channels in bacterial outer-membranes are small trimeric β barrels inserted via the BAM complex (Selkrig *et al.*, 2014) (Introduction Section 1.3.4). There are limitations to the size of substrates which can be exported through these small complexes and as such Bacteria have developed a larger pore, the secretin, which allows for the passage of large substrates (Collin *et al.*, 2011).

The secretin is found in most diderm bacteria and used to facilitate trafficking of large proteins. This pore is common to several systems including the T2SS (Nivaskumar and Francetic, 2014), the injectisome T3SS (Sha *et al.*, 2004), T4P (Pelicic, 2008), the DNA uptake systems (Knapp *et al.*, 2017), and the filamentous bacteriophage extrusion system (Russel, 1994). In T3SS the pore facilitates the passage of the needle complex through the outer membrane to allow transfer of effector molecules to a target cell. The T4P are homologous to the T2SS and share a similar structure. The difference is that the pilus in the T4P extends through the outer membrane to interact with surfaces.

1.3.3.4 Type V secretion

Type V secretion systems (TVSS), or autotransporters, were previously thought to be an autonomous system for targeting proteins to the OM. For a review please see (Grijpstra *et al.*, 2013). These proteins consist of a ß-barrel and a passenger domain. The ß-barrel incorporates into the OM, while the passenger domain can be cleaved and released to the extracellular media or remain tethered to the bacterium. Passenger domains have been found with many activities, including adhesion and enzymatic activity (van Ulsen *et al.*, 2014). These systems are assembled by the ß-barrel assembly complexes discussed in the next section.

1.3.4 ß-Barrel Assembly

Of all the systems found in the OM of diderm bacteria, one of the most important and fundamental is the integration of proteins within the OM. Unlike CM proteins, OM proteins cannot use the SEC pathway for OM insertion as the periplasmic space lacks the necessary ATP. Within diderm bacteria, the integral proteins of the OM have a ß-barrel fold that spans the OM (Knowles *et al.*, 2009). In contrast to this, all integral proteins of the CM, as well as the CM of Archeae and Eukaryotes have an α-helix structure and are inserted via SEC (Zückert, 2014). The structure of these ß-barrels usually consists of antiparallel strands from 8-26, and most functional porins, those with a central pore, have 16-18 strands. These proteins are assembled by five known, homologous systems (Figure I.13): the ß-barrel assembly apparatus (BAM), translocation and assembly module (TAM), the two-partner secretion system (TPS), the toc75 system of plastids, and the sorting and assembly machinery (SAM) of mitochondria. These systems are referred to as Omp85 proteins as the key assembly protein from each belongs to the Omp85 protein family (BamA, TamA, TpsB, Toc75-III-IV, and Sam50) (Voulhoux *et al.*, 2003). The SAM and Toc75 systems are highly divergent from Bacteria and as such will not be discussed here, however, a recent review can be found here (Lee *et al.*, 2014).

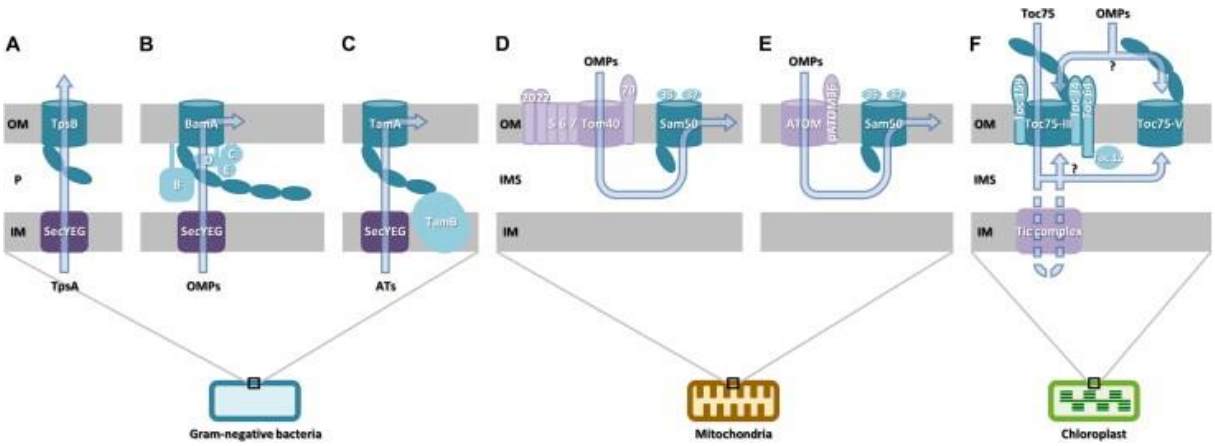

Figure I.13. Mechanisms of ß-Barrel Assembly.
Showing bacterial systems TPS (A), BAM (B), TAM (C); mitochondrial system SAM (D & E); and plastid system (F). Taken with permission from (Jacob-Dubuisson *et al.*, 2013).

1.3.4.1 BAM

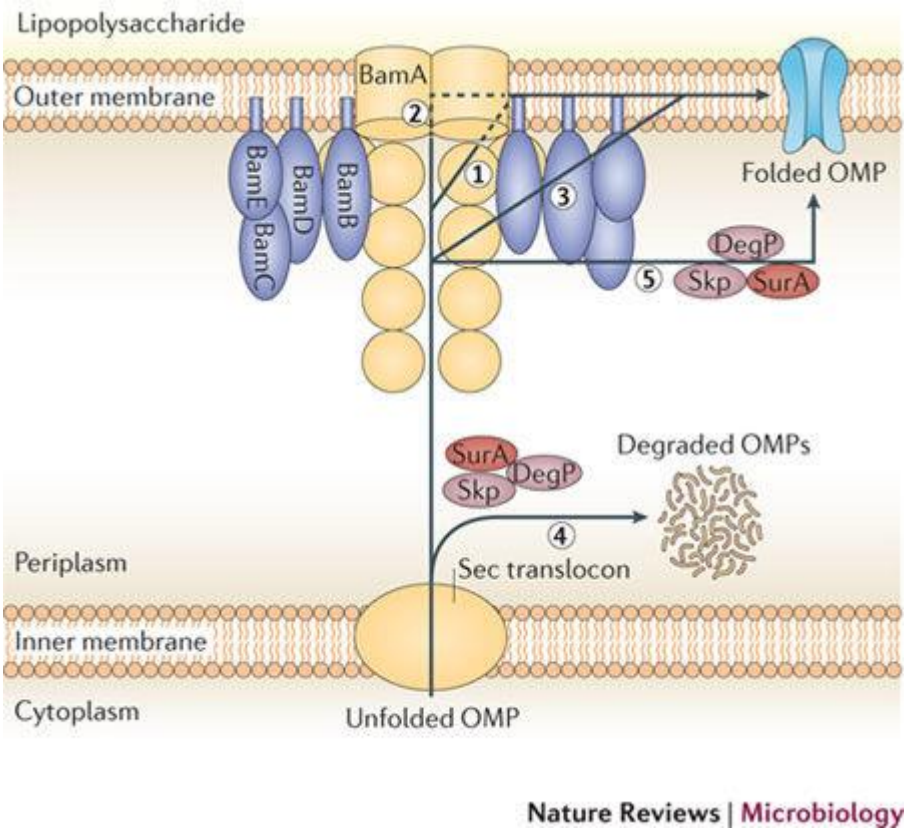

Figure I.14. BAM System of *E.coli*.
Proteins destined to be inserted by the BAM system are first translocated past the CM by the SEC translocase, then carried by SurA or Skp to the OM, and finally inserted and folded by the Omp85 protein BamA and the accessory proteins BamB, C, D, E. Taken with permission from (Leyton *et al.*, 2012).

The first system to be discovered, and the best described, is the BAM system (Figure I.14) has been predominantly studied in *E. coli*. For a recent review please see (Noinaj *et al.*, 2017). To be assembled the OM proteins must first be translocated past the CM via the SEC pathway. These proteins are then transported through the periplasm to the OM, which is completed via two different chaperones: SurA and Skp. SurA has been shown to be the primary chaperone responsible for transferring the majority of the proteins to the OM while Skp forms a rescue function to aid SurA (Volokhina *et al.*, 2011). Should both these proteins fail, or the final protein becomes misfolded, DegP will degrade the protein and recycle the amino acids. The two chaperones transport the unfolded protein to the five component BAM complex, consisting of BamA-E. The Omp85 protein BamA, an OM protein, is an essential part of the complex and performs the necessary integration of the protein to be folded (Albrecht *et al.*, 2014). In addition to the Omp85 domain, BamA contains POTRA domains which function as another co-factor necessary for stable integration of proteins into the OM (Simmerman *et al.*, 2014). The other proteins of the complex, BamB,C, E, are all non-essential lipoproteins which aid BamA in its function (Noinaj *et al.*, 2011; Warner *et al.*, 2011; Jeeves *et al.*, 2015) except BamD which has been found to be an essential lipoprotein (Rossiter *et al.*, 2011). The most important of these proteins, BamB, interacts closely with BamA while the remainder forms a complex which interacts with BamB.

1.3.4.2 TAM

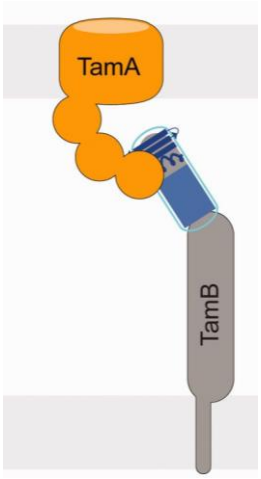

Figure I.15. TAM System of *E.coli*.
The TAM complex assembles proteins in the OM with only two proteins: the CM tethered chaperone TamB and the OM Omp85 TamA. Modified with permission from (Heinz et al., 2015).

Within *E. coli* a second system of OM ß-barrel was discovered, TAM (Figure I.15) (Selkrig *et al.*, 2012). This system contains only two proteins; TamA, an Omp85 protein

found in the OM, and the chaperone TamB. TamB is integrated into the CM and spans the periplasmic space to interact with TamA (Iqbal *et al.*, 2016). No other co-factors nor mechanism have been identified, however, work on the TAM system is still in its infancy. This two protein complex has been shown to be responsible for insertion of many proteins, not least of which are the autotransporters. It was recently found that many autotransporters are in fact assembled using the TAM complex (Selkrig *et al.*, 2012), although others are assembled via BAM (Grijpstra *et al.*, 2013; van Ulsen *et al.*, 2014).

1.3.4.1 TPS

The TPS system is a component of the TVSS. They are composed of an OM TpsB Omp85 transporter and a TpsA cargo protein. Both components are first transported past the CM via the SEC pathway and then the TpsA cargo is transported through the OM exclusively via the Omp85 transporter. Although little is known about the mechanism of this system some details have been determined. A recent review can be found here (Jacob-Dubuisson *et al.*, 2013).

1.3.4.2 Atypical systems and Evolution

The distinct nature of BAM and TAM for assembling ß-barrel proteins was challenged through a close look at Borrelia ß-barrel assembly. In *Borrelia burgdorferi* BamA, B, D, Skp and TamB were found to work in concert to produce functional proteins in the OM (Iqbal *et al.*, 2016). This hybrid system is curious and begs the question "which components of the ß-barrel assembly system are ancestral?" Since then two papers have focused on the evolution of the BAM (Webb *et al.*, 2012) and the TAM (Heinz *et al.*, 2015). The Webb paper focuses on the evolution by means of presence/absence of BamB, C, D, E and POTRA domains within BamA. They found that only the γ-Proteobacteria possess a complete system and most cofactors are limited in distribution to the Proteobacteria. Notable exceptions were the *B. burgdorferi* proteins previously mentioned and BamD in the Bacteroidetes. The paper focusing on TAM evolution used more diverse and complete methods (Heinz *et al.*, 2015). They discovered that the true two component TAM system was exclusively confined to the Proteobacteria and that the ancestor of all ß-barrel assembly is a two-component system with BamA and TamB that later duplicated and acquired co-factors.

1.3.5 Lipoproteins

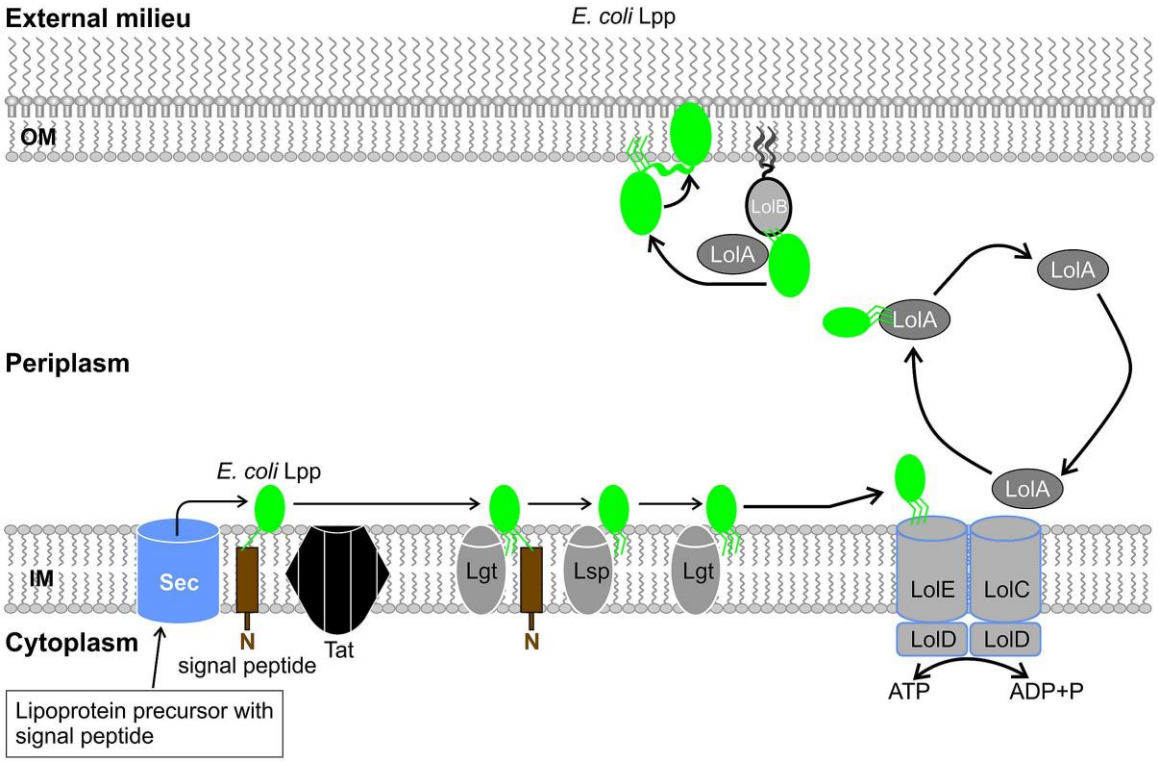

Figure I.16. Lipoprotein Modification and Transport to the OM.
Bacterial lipoproteins are first transported via sec past the CM. They are then modified by Lgt, Lsp, and Lnt, after which they are transferred to the shuttle protein LolA with the energy obtained from the ATPase complex LolC, D, E. Finally. the lipoprotein is transferred from LolA to LolB and then integrated into the OM. Taken with permission from (Goolab et al., 2015)

Post-translational lipid modification of proteins is a fundamental component of bacterial physiology. Lipid modification allows proteins to be tethered to a membrane and associated locally. For a recent review, please see (Buddelmeijer, 2015). This process is used by many proteins, with 1-3% being lipid modified (Nakayama *et al.*, 2012). Proteins are targeted to be lipoproteins through a specific signal sequence known as a lipobox. First, the proteins have a diacylglycerol group added by the Lgt enzyme. Then Lsp cleaves off the signal sequence. Finally, they are acylated by Lnt. Lgt and Lsp are universally distributed among Bacteria, however, Lnt has only been found in Proteobacteria and Actinomycetes. This process allows lipoproteins to be incorporated into the CM, however, transfer to the OM requires additional equipment (Figure I.16).

After the lipidation by Lgt, Lsp, and Lnt, the Lol system is necessary for diderm bacteria to traffic lipidated proteins for attachment to the OM. The system itself contains three main components: an OM receptor (LolB), a periplasmic carrier protein (LolA), and an

ABC transporter complex (LolCDE). The proteins are synthesized in the cytoplasm and lipidated on the CM. Following the lipidation, they are transported via the sec pathway to the outer leaflet of the CM. These proteins will stay on this leaflet unless they possess the proper signal sequence within the first few amino acids. The lipoproteins bind the LolCDE complex which transfers the lipoprotein to LolA through ATP hydrolysis. Following this LolA passes the lipoprotein to LolB, which finally releases the protein to the inner leaflet of the OM (Zückert, 2014). Most components of this pathway are well conserved throughout diderm bacteria, however, LolB is restricted to the Proteobacteria (Sutcliffe *et al.*, 2012).

1.3.6 TonB dependent transport

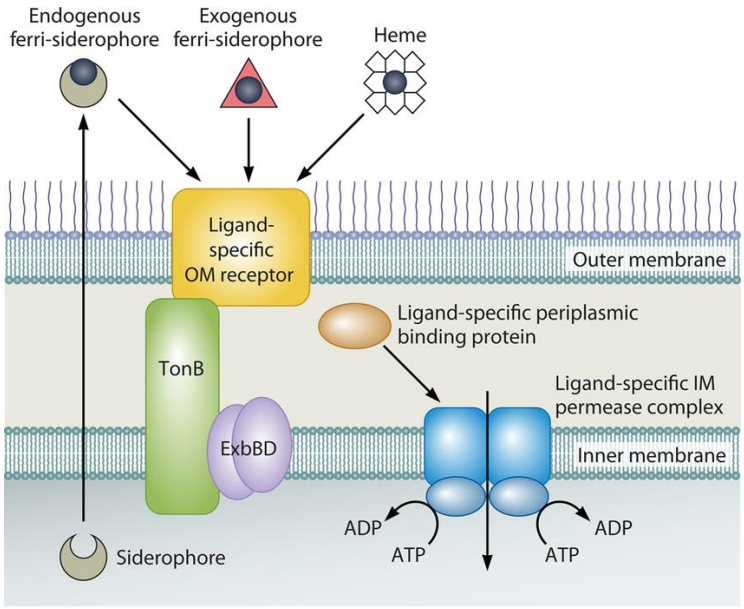

Figure I.17. TonB Dependent Transport.
In uptake of molecules by the TonB system, first, a carrier is secreted that binds to the desired ligand. This ligand/carrier complex is then transported through a TonB receptor that acquires its energy through the ExbBD/TonB ATPase complex. It is transferred to a binding protein and shuttled into the cytoplasm via an ABC transporter.

The OM prevents the entry of toxins and other problematic molecules, however, it also prevents the entry of necessary molecules such as sugar or iron. As such, diderm bacteria have other systems for dedicated active transport of key molecules; one of the most important systems is the TonB transport system (Krewulak and Vogel, 2011) (Figure I.17). Bacteria can have multiple different copies of this system within their genome, with each being specific to a different molecule, with most work focused on iron. This system grabs iron by first secreting a siderophore (other molecules use other chelators). The siderophore-Fe complex

then binds to the TonB receptor in the OM. As this is active transport the system needs to be energized for the siderophore-Fe molecule to pass through the OM. This is accomplished through TonB, ExbB, and ExbD, which can hydrolyze ATP and pass it to the OM receptors. After passing through the membrane the siderophore-Fe complex binds to a periplasmic binding protein and passes the CM via an ABC transporter. Although ABC transporters are found in both diderm and monoderm bacteria it is worth noting that periplasmic binding proteins of monoderm bacteria are usually lipidated or contain PG binding domains, while in diderm bacteria they are normally completely soluble.

1.3.7 Outer Membrane Attachment

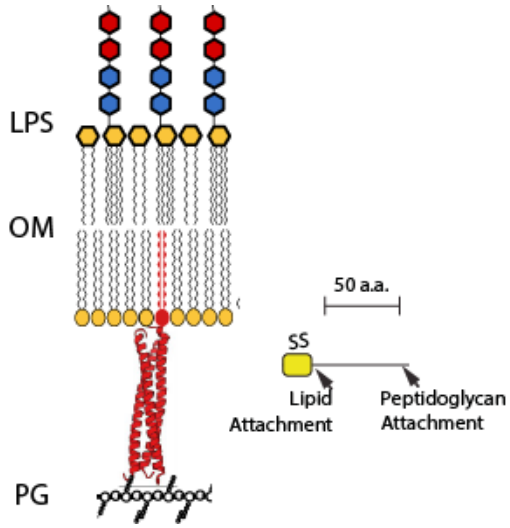

Figure I.18. Lpp.
Structure of the major OM tether of *E. coli*, Lpp ((PDB ID: 1EQ7) (Left) and sequence characteristics (Right)

Diderm bacteria have a unique requirement not found in organisms with a single membrane; they must tether their OM to the PG. This tethering method was discovered in *E. coli* by Braun and colleagues in the late 1960's (Braun and Rehn, 1969). This protein, Lpp or Braun's Lipoprotein (Figure I.18), is the most abundant protein at roughly 1 million copies per cell and it tethers the OM via a lipid moiety which integrates into the lipid bilayer (Guo *et al.*, 2014). The small, 58 amino acids, molecule forms a trimer in vivo with one of the three molecules covalently bound to the diaminopimelate residue in the peptide cross bridge of PG (Shu *et al.*, 2000). Experimental evidence has been found for homologs in *Pseudomonas aeruginosa*, *E. coli*, *Salmonella enterocolitica*, and other γ-Proteobacteria (Ching and Inouye, 1985).

1.4 Monoderm Characteristics

There are no defining markers that are found in all monoderm bacteria and absent from diderm bacteria, however, there are a few that are present in most monoderms and absent from almost all diderms. I will present the best-known systems here both the PG attachment mechanism of teichoic acids and the sortases (Silhavy *et al.*, 2010).

1.4.1 Teichoic Acids and Lipoteichoic acids

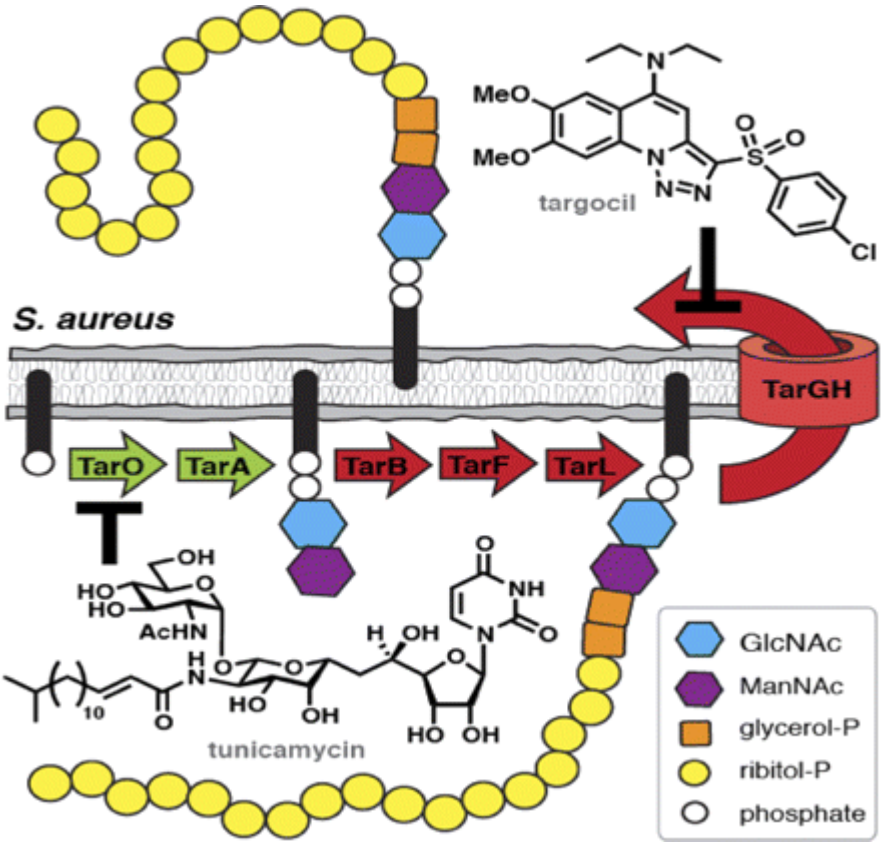

Figure I.19. Wall Teichoic Acid Biosynthesis.
The teichoic acids are synthesized in the cytoplasm through lipidation of a saccharide by TarO. The saccharide is then dimerized by TarA. After which additional saccharides and ribitol are added by TarB, F, L. The molecule is then flipped past the CM by the TarGH translocase. Taken with permission from (Campbell *et al.*, 2012)

The PG sacculus of monoderm bacteria is a thick mesh of proteins and polysaccharides in addition to the actual molecule of PG. Within this mesh are teichoic acids; a diverse family of glycopolymers containing phosphodiester-linked polyol repeat units. Two main families exist, the membrane anchored lipoteichoic acids (LTA) and the wall teichoic acids (WTA), which are covalently attached to the peptidoglycan. WTA's, reviewed in (Brown *et al.*, 2013), are among the most abundant PG-linked polymers in most monoderm bacteria and are important for maintaining cell shape. Furthermore, they are required for

antibiotic resistance in many strains. WTA's have an incredible diversity in structure and the only commonality among all of them is the linkage unit that is attached to the PG, a trisaccharide and the biosynthesis mechanism (Figure I.19).

LTA's, broadly speaking, are polymeric chains of poly-hydroxy alkane anchored to the envelope through a lipid. Five types of LTA's have been classified based on the type of polymer, although the most common type is a polymer of 1,2-polyglycerol phosphate. All other types are deviations on this common theme and have a synthesis pathway that is completely distinct from that of WTAs (Schneewind and Missiakas, 2014).

1.4.2 Sortases

Sortases are a group of enzymes mostly exclusive to monoderm bacteria that are responsible for the covalent attachment of proteins to the PG exoskeleton in order to be displayed on the surface of the bacterium (Spirig *et al.*, 2011). Sortases are found in most monoderm bacteria, a few diderm bacteria, and some Archeae. Proteins displayed include virulence factors and enzymes similar to those displayed by SLH proteins or autotransporters. The archetypal sortase recognizes any protein with a carboxyl terminal sorting signal of LPXTG. The protein is cleaved between the threonine and the glycine, and then it is covalently attached to the PG via transpeptidation (Dramsi *et al.*, 2008; Spirig *et al.*, 2011; Schneewind and Missiakas, 2012).

2 Current Theories on the Evolutionary History of the Bacterial Cell Envelope

Over the years several theories have been suggested as to how the bacterial envelope originated. These theories are divided into two classes: diderm first, where the last common ancestor (LCA) of Bacteria was a diderm, and monoderm first, where the LCA of Bacteria was a monoderm.

2.1 The Blobel Hypothesis

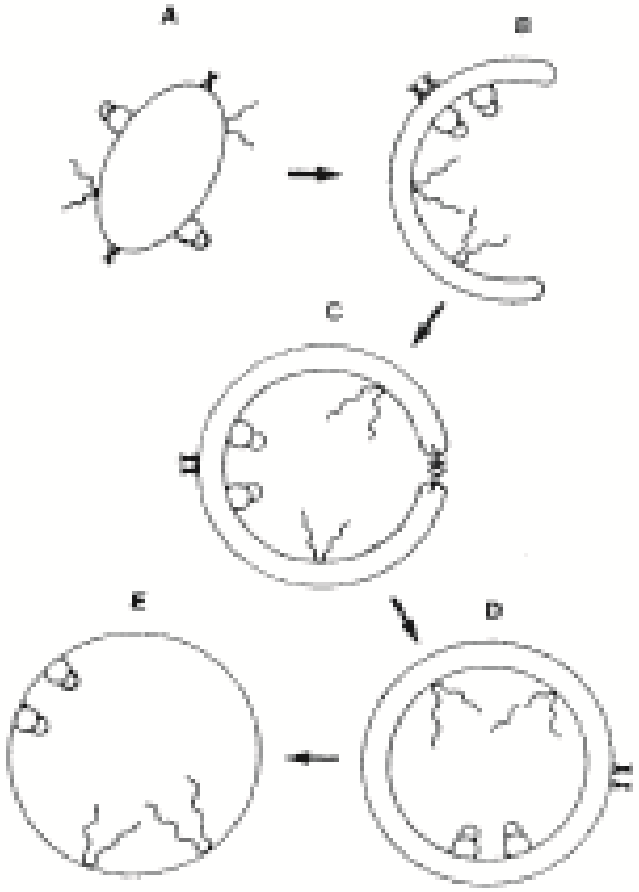

Figure I.20. The Blobel Hypothesis.
In the time of primordial cells (A), cellular systems were attached to the outside of a hydrophobic sphere. These systems developed complexity and the lipid thinned (B). As organization progressed the lipid invaginated and formed a cell like structure (C). Eventually, the invagination closed on itself and produced the last common ancestor of Bacteria (D). Monoderm clades then immediately lost the OM (E). Taken with permission from (Blobel, 1980)

The earliest hypothesis on the origin of bacterial cell envelopes comes from Nobel Laureate Gunter Blobel in his 1980 paper regarding the origin of the signal sequence and how proteins originally crossed the membrane (Blobel, 1980). He proposed that cellular systems originated on the exterior surface of the last universal common ancestor (LUCA) (Figure I.20). These systems were attached to a vesicle by hydrophobic interactions. As the proteins gained the capability to integrate into the membrane the vesicle began to invaginate and fold in on itself. After this folding, the first cellular organism was left with a double membrane. This theory not only suggests a diderm first origin of Bacteria but of all cellular life. This paper was based on logic and thought and has no concern for diversity or to phylogeny itself. Nonetheless, it is still an interesting hypothesis worthy of consideration.

2.2 Cavalier-Smith Hypothesis

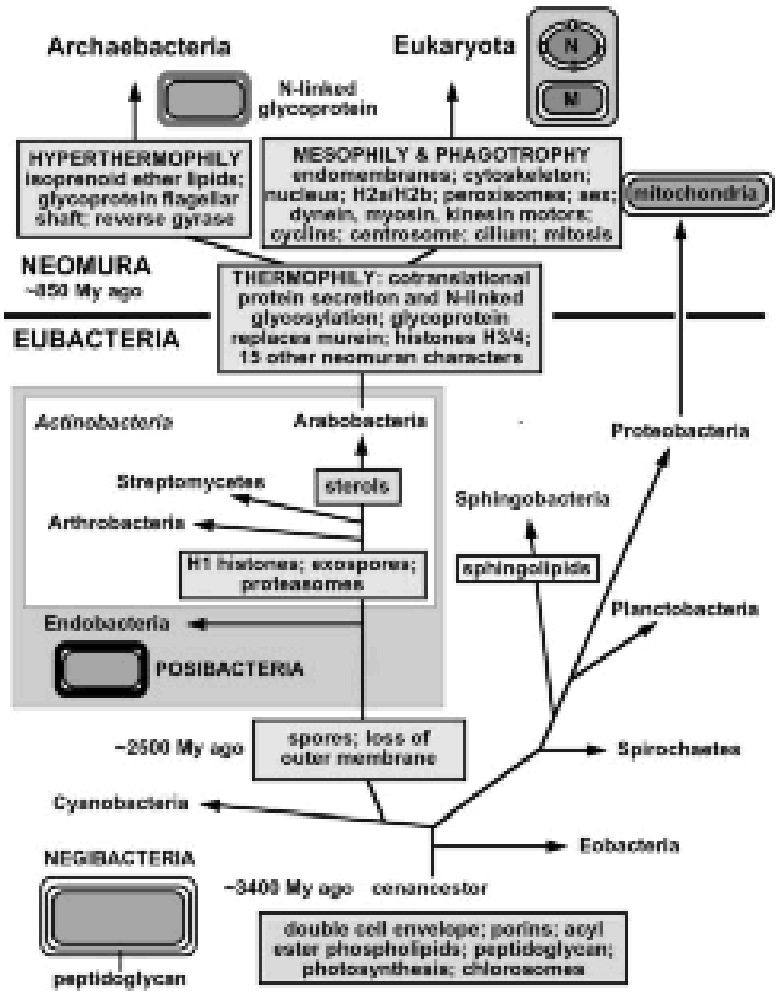

Figure I.21. The Cavalier-Smith Hypothesis.
The first cells were diderm bacteria (Negibacteria in the Figure with the description in bottom box), these then diverged into modern diderm bacteria (the right branch) and monoderms (Left branch, Posibacteria in Figure). The diderms had many branching lineages with the most recently diverging being the Proteobacteria. Before the left branch lost the OM the Cyanobacteria diverged and the monoderms diversified and evolved into Eukaryotes and Archea. Taken with permission from (Cavalier-Smith, 2002)

In a massive 83 page manuscript by Cavalier-Smith (Cavalier-Smith, 2002), a diderm first view that extends from the Blobel hypothesis is presented. This work places the origin of cellular life in the diderm bacteria based on transition analysis. He suggests that the diderm bacteria lost the OM in a single event to form all monoderm bacteria. From this single loss, it is suggested that all monoderm Bacteria, Eukaryotes, and Archea evolved from a sole diderm ancestor (Figure I.21).

This grand hypothesis is presented as a step by step evolution. It is claimed that one of the earliest inventions in life was the flagella after the Cyanobacteria and Chloroflexi had

emerged. It is then claimed that all other diderm bacteria must group with the Cyanobacteria and Chlorflexi, due to the presence of Omp85. Unfortunately, these careful logical steps have since been disproven; mostly due to the determination that the Chloroflexi are monoderms (Sutcliffe, 2011). This breaks the theory of a single origin of monoderm life, however many aspects of the theory may still be true.

2.3 The Gupta Hypothesis

One of the earliest presentations for the theories of monoderm first evolution comes from Dr Gupta. The most relevant topic to this discussion is that of Firmicutes, where it is stated that antibiotic selection led to the development of the OM in monoderm bacteria (Gupta, 2011). This theory (Figure I..22) states that monoderm bacteria developed into diderms in at least three instances, one for the Thermotoga, one for the atypical diderms, and a third for the traditional diderms such as Proteobacteria (Figure I.22). This hypothesis is based on the presence of insertions within the conserved heat shock proteins 60 and 70.

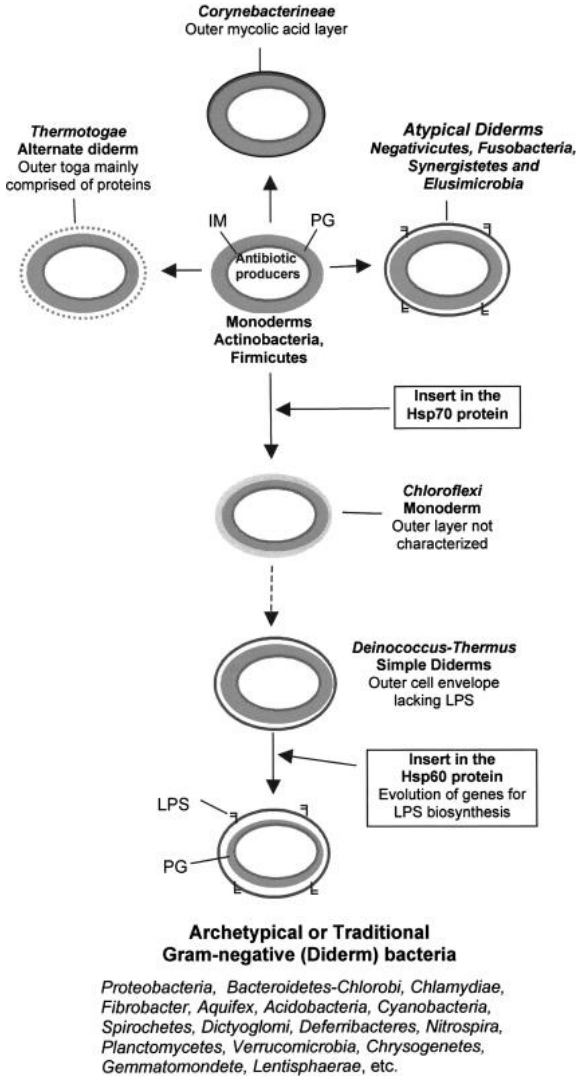

Figure I.22. The Gupta Hypothesis.
The first Bacteria were monoderm. The cells then diverged into at least 4 different types of envelopes that had different OMs with additional intermediate steps representing other lineages. Obtained from (Gupta, 2011) with permission.

Gupta goes on to hypothesize on the selection pressure for this transition from monoderm to diderm, giving the following four arguments:

1) Monoderm bacteria are the dominant producer of antibiotics
2) Antibiotic production gives an advantage over non-producing bacteria
3) Resistance to antibiotics can develop by a variety of mechanisms
4) Monoderm bacteria have a higher sensitivity to antibiotics than diderms

Although the original evidence does not hold up to modern phylogenetic methods, this logic of selection pressure is still valid and may have happened in some bacterial clades.

2.4 The Lake Hypothesis

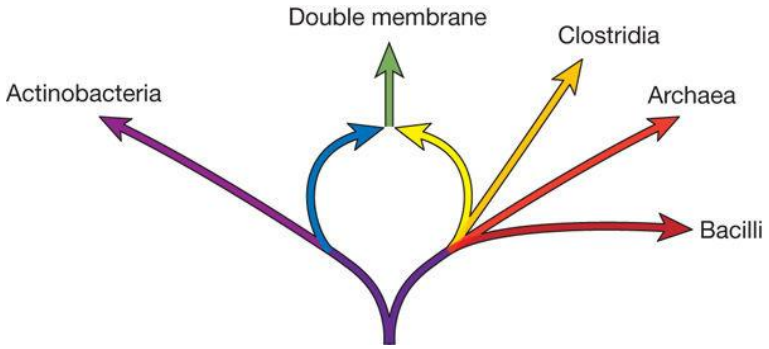

Figure I.23. The Lake Hypothesis.
The first Bacteria were monoderm. Through a bacterial fusion of two monoderm groups, the Actinobacteria and Clostridia, a diderm bacteria was formed. Taken with permission from (Lake, 2009)

Lake proposed an alternative and controversial theory for the origin of the bacterial OM (Lake, 2009) (Figure I.23). This theory proposed that the OM is the result of endosymbiosis of one bacterial clade into another, in this case, Actinobacteria and Firmicutes. This hypothesis was constructed by searching for protein families that were transferred from one ancestral group and passed to another. By using roughly 500 protein families it was determined that an actinobacterium was endosymbiosed by a clostridium. This endosymbiosis produced a double membraned bacterium, which diverged into all modern diderm lineages.

Unfortunately, this article is plagued with artefacts and has been extensively criticized and reanalysed (Swithers *et al.*, 2011). The major fault in Lake's analysis is the assumption that all diderm clades are monophyletic. This assumption has no basis in the literature or definitive proof. The taxa selection was quite small, with only 37 taxa to cover the diversity of all Bacteria and Archaea. Finally, the selection was biased as only small portion of available PFAM domains was used and the selection criteria were not clear. Due to the complete debunking of this theory and lack of any support, it will not be considered in any regard for the remainder of the thesis.

2.5 Errington or L-form Hypothesis

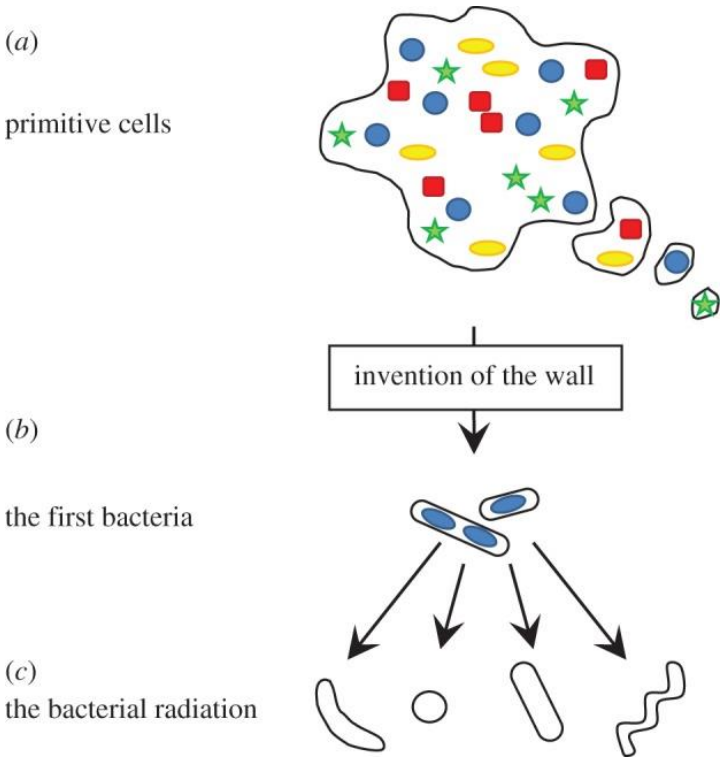

Figure I.24. The Errington Hypothesis.
The first cells were a primitive sack of cellular systems that replicated by budding off a small portion of the cellular mass (A). Then PG was invented allowing the first true Bacteria, a monoderm (B). This form diverged into all Bacteria we know today (C). Taken with permission from (Errington *et al.*, 2016)

One of the most intriguing works of recent years is the L-form bacteria or bacteria without PG. L-form bacteria are rarely formed in nature with only a single clade, the Tenericutes, lacking any form of peptidoglycan. The Tenericutes are well known as they include the *Mycoplasma* (Whitman, 2011). The Errington laboratory has had great success transforming *B. subtilis* into an L-form by deletion of the DCW cluster (Leaver *et al.*, 2009). L-form bacteria reproduce by budding and have a much simpler cellular structure than another organism. Errington recently proposed that all life started out as an L-form life form (Errington *et al.*, 2016) without any major cellular structure (Figure I.24). This simple life form would not require any proteinaceous machinery for division or replication and may be the simplest possible division mechanism. He does not go into depth, but this recent hypothesis suggests a monoderm first origin of Bacteria as no true diderm L-form bacteria have been produced (Joseleau-Petit *et al.*, 2007).

3 Distribution of Bacterial Envelope Architectures

3.1 Phylum Level Cell Envelope Architecture

As stated before, the bacterial cell envelope is stereotypically defined into diderm and monoderm with *E.coli* representing the former and *B. subtilis* representing the latter. Although this is the textbook example of cell envelope architecture it does not account for the massive bacterial diversity available and any hope of understanding the evolution of envelopes requires breaking the stereotype. At the beginning of my thesis, the best representative of bacterial diversity was a review by Dr Sutcliffe (Sutcliffe, 2010) (Table I.1). This table summarizes the entire article, where he compiled all available literature for known diderms with a particular focus on isolated LPS or envelope structure. He then queried available genomes for LpxC and KdsA to determine if each phylum had the potential for producing LPS. The data suggests a wider distribution of LPS producing genes than previously thought from some of the most diverse clades. Both the Cyanobacteria and the Proteobacteria have described LPS, however, the evolutionary origin of these genes has not been fully described. It has been suggested that the precise inheritance of LPS biosynthesis genes is vertical (Opiyo *et al.*, 2010). Aside from this, some clades with atypical envelopes do exist. The Thermotoga have a slightly different OM components; they lack LPS yet still produce β-barrels in their OM (Petrus *et al.*, 2012). Another interesting clade is the PVC, which appears to lack a PG sacculus, yet still encode the genes necessary for biosynthesis (Jacquier *et al.*, 2015). Apart from these unique cases, the OM of the diderm phyla are remarkably similar and may have a common origin.

Aside from the diderm groups, there are three clear monoderm phyla; Actinobacteria, Firmicutes, and Chloroflexi (Table I.1); with the remainder being diderm or unknown. One of the classically regarded monoderm groups is the Actinobacteria, which surprisingly contains diderm members. These diderm members are the Corynebacteriales, which have a unique type of OM not found elsewhere in life. This OM is comprised of Arabinogalactan covalently attached to the PG with a hydrophobic layer of mycolic acids on the exterior (Bansal-Mutalik and Nikaido, 2014). This envelope lacks LPS and β-barrels and is believed to be a recent development within the clade, a *de nova* membrane (Jamet *et al.*, 2015). Unfortunately, this does not tell us if the ancestor of Bacteria was a diderm or monoderm as the Actinobacteria may have lost their OM and developed a new one later. In contrast, little is known of the Chloroflexi envelope structure. It was previously thought to be a diderm phylum, however

high-quality EM has determined it to be a monoderm (Sutcliffe, 2011). The final monoderm phylum, the Firmicutes, is discussed in depth in the next section.

Phylum	Representative genera	Characterised LPS	LPS biosynthetic genes detected [b]	Refs
Acidobacteria	Acidobacterium , Solibacter	No reported isolation	Present in 3/3 genomes	[27]
Aquificae	Aquifex	Present	Present in 5/5 genomes	[51,66]
Bacteroidetes	Bacteroides , Porphyromonas	Present	Present in 7/7 genomes	[67,68]
Chlamydiae	Chlamydia	Present	Present in 5/5 genomes	[69]
Chlorobi	Chlorobium	Present	Present in 10/10 genomes	[70]
Chloroflexi	Chloroflexis , Herpetisphon , Dehalococcoides , Thermomicrobium	Absent	Absent from 11/11 genomes	[57–59]
Chrysiogenetes	Chrysiogenes	No reported isolation	No genome available	
Cyanobacteria	Prochlorococcus , Synecococcus	Present but KDO absent in some taxa	Present in 8/8 genomes; some variation in distribution of KdsA homologues	[71]
Deferribacteres	Deferribacter , Flexistipes , Denitrovibrio	No reported isolation	Present in 1/1 genome	
Deinococcales –Thermus	Deinococcus , Thermus	Absent	Absent[c]	[42,45,46]
Dictyoglomi	Dictyoglomus	No reported isolation; complex intercellular cell envelope structures evident	LpxC present in 2/2 genomes; no clear orthologues of KdsA	[72,73]
Elusimicrobia	"Elusimicrobium minutum"	No reported isolation	Present in 1/1 genome	[74]
Fibrobacteres	Fibrobacter	Absent	No clear orthologues present	[62]
Fusobacteria	Fusobacterium	Present	Present in 10/10 genome	[75]
Gemmatimonadetes	Gemmatimonas	No reported isolation	Present in 1/1 genome	
Ktedonobacteria	Ktedonobacter [d]	No reported isolation	No genome available	
Lentisphaerae	Lentisphaera , Victivallis	No reported isolation	Present in 1/1 Lentisphaera genome; LpxC apparently absent in Victivallis vadensis	[27]
Nitrospira	Nitrospira , Thermodesulfovibrio , Leptospirillum	Present	Present in 1/1 genome	[76]
Planctomycetes Verrucomicrobia	Planctomyces Verrucomicrobium , Gemmata , Pirullela	Present	Present in 11/11 genomes	[27,77,78]
Spirochaetes	Leptospira , Treponema	Present in some genera only	Present in 4/4 genomes for Brachyspira and Leptospira ; Absent from 12/12 genomes for Borellia and Treponema	[34,35,79]
Synergistetes	Anaerobaculum , Dethiosulfovibrio , Thermanaerovibrio	No reported isolation	Present in 3/3 genomes	[80]
Thermodesulfobacteria	Thermodesulfobacterium	No fully characterized representative [e]	No genome available	[65]
Thermotogae	Thermotoga	Absent	LpxC absent from 10/10 genomes; distant homologues of KdsA detected in 3/10 genomes	[51]

[a]Only phyla with cultivated representatives, excluding the Actinobacteria , Firmicutes and Proteobacteria , were included.
[b]Homologues were identified by unfiltered BLASTP searches of a representative selection of microbial genomes available at http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?organism=microb . The E. coli K-12 LpxB, LpxC and KdsA sequences were used as queries and searches were performed with low stringency (cut-off, E=0.01).
[c]LpxC homologues were identified in 2/2 Meiothermus genomes.
[d]Originally described as ' Ktedobacter '. Status as a phylum unclear; might be a class within Chloroflexi [21,23] .
[e]See text.

Table I.1. The Diversity of Bacterial Cell Envelopes across Bacterial Phyla.
Table showing the major bacterial phyla and whether they have LPS characterized. Taken with permission from (Sutcliffe, 2010)

In an ideal world, we would resolve the phylogeny of all Bacteria and compare this phylogeny to that of OM proteins and determine if the OM was gained, transferred, or lost. Unfortunately, the Bacterial phylogeny is not yet fully resolved (Wu *et al.*, 2009) and OM genes do not usually give the necessary resolution as well (Opiyo *et al.*, 2010). Fortunately, we have an ideal phylum that contains both monoderm and diderm members; The Firmicutes.

4 Firmicutes: The perfect model clade

As there is no confidence on the envelope structure of the LCA of Bacteria, we need a model clade which can be studied to provide evidence for the question at large. Allow me to introduce the Firmicutes, one of the best studied Bacterial phyla.

Firmicutes were originally described to encompass all Bacteria which had a Gram Positive cell structure, this phylum was later split into two groups with the Actinobacteria being high GC% Gram-positive and the Firmicutes as low GC% Gram Positives (Whitman, 2011). These names, like Gram-positive itself, have become antiquated as there are many exceptions in both groups when it comes to naming. The Firmicutes are one of the oldest and best studied of all the Bacterial phyla with an entire Volume of Bergey's Manual of Systematic Bacteriology dedicated to their discussion (Whitman, 2011). They also encompass a broad range of metabolisms, morphologies, and habitats. As such, very few characteristics are exclusive to this group, with the exception of sporulation; classical sporulation, exemplified by *B. subtilis* and *C. difficile* is a characteristic of many Firmicutes not found in other bacterial phyla (Tocheva *et al.*, 2016).

Even their defining characteristic of monoderm envelopes has been broken, for we have the unexpected. The Firmicutes include two clades that possess an OM: the Negativicutes and the Halanaerobiales.

4.1 Negativicutes

Although the Negativicutes were first observed in the 17th century, the first modern description of a member, *V. parvula*, came in 1898 by Veillon and Zuber in their landmark paper on anaerobic bacteria (Veillon and Zuber, 1898). The key defining feature of this clade is that they are diderm. It was never thought that these bacteria were Firmicutes as, by definition, they were diderms. It was not known that they were Firmicutes until 16S rRNA sequencing was available. Before this, many species were placed within the Proteobacteria or the Bacteroides over the years based on their biochemical and physical characteristics (Whitman, 2011).

Negativicutes have been proposed to form a Class within Clostridia (Marchandin *et al.*, 2010). The Negativicutes have been divided into four families (Figure I.25) (Campbell *et al.*, 2015) and currently contains 28 described genera (Euzeby, 1997). Most Negativicutes are anaerobic and live in moderate temperature environments. Some sporulate, and some are present in the human body.

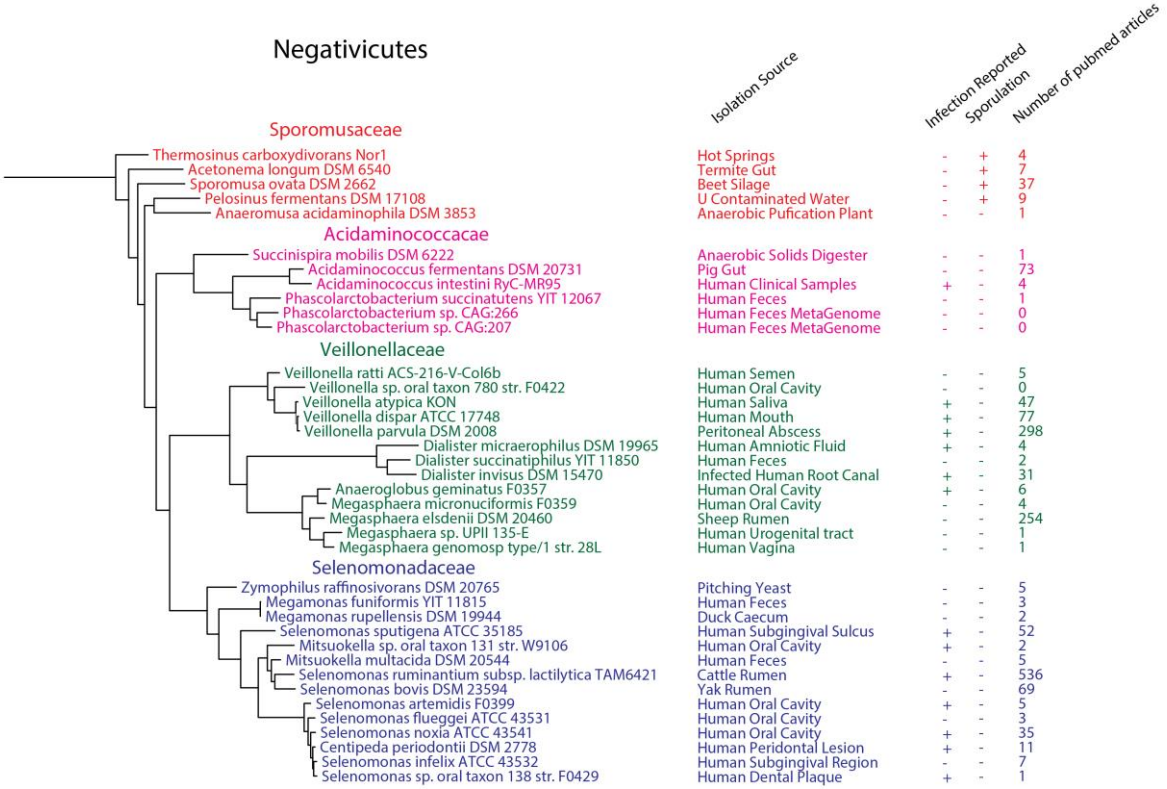

Figure I.25. Negativicute Genomes Available as of October 2014.
Four major families have been assigned (unique colours) and the strain details and number of PubMed articles are listed. Phylogeny was taken from (Antunes *et al.*, 2016)

4.1.1 Sporomusaceae

The deep emerging branch of the Negativicutes, the Sporomusaceae, has many unique characteristics not found elsewhere in life. The Sporomusaceae are the only Negativicutes that are known to sporulate, with a well-studied pathway in *A. longum* (Introduction Section 4.4.2). While the remainder of the Negativicutes are usually mammal or bird associated, the Sporomusaceae inhabit a wide variety of environmental habitats. For instance, the *Thermosinus* inhabit hot springs (Sokolova *et al.*, 2004), the *Pelosinus* live in heavy metal contaminated ground water (De León *et al.*, 2015), and *Acetonema*'s niche is the hind-gut of Termites (Kane and Breznak, 1991). With a wide range of habitats, it is no surprise that the known Sporomusaceae have a diverse array of metabolism, morphology, and physiology (Whitman, 2011). Due to the environmental niche of these bacteria, there is the greatest potential for discovery of new species and genomes. This is due to the fact that they are not part of the human microbiome project or any of the other mammal sequencing efforts, unlike the other Negativicute Families (Turnbaugh *et al.*, 2007; Aujoulat *et al.*, 2014).

4.1.2 Selenomonadaceae

The Selenomonadaceae have a much narrower habitat than that of the Sporomusaceae, being that they predominantly inhabit cattle rumen and the human mouth, although *Megamonoas rupellensis* has been isolated from the caecum of a duck (Whitman, 2011). The group has predominantly been studied as a normal component of cattle rumen (especially *Selenomonas ruminantium*), for how it affects milk, digestion, and methane production (Yang *et al.*, 2016). Aside from being a cattle commensal, it can be a human pathogen responsible for bacteraemia (Pomeroy *et al.*, 1987) and periodontitis (Tanner *et al.*, 1989). An interesting feature of many Selenomonadacae is the presence of a lateral helical flagellum, allowing a "corkscrew" motility (Males *et al.*, 1984; Zhang and Dong, 2009).

4.1.3 Acidaminococcacae

The Acidaminococcae is one of the least diverse families of the Negativicutes, with only 4 currently described genera and the dominant research focusing on *Acidaminococcus*. Along with a restricted diversity, they also are limited in niche as they are only found in human and porcine digestive tracts. Within these niches, they represent a strong portion of the environment with *Acidaminococcus intestinalis* comprising around 1% of the porcine faecal

bacterial population (Callaway *et al.*, 2010). Furthermore, they have been implicated as an antibiotic resistance reservoir (Galán *et al.*, 2000) and can cause human infection in rare cases (Marchandin *et al.*, 2010).

4.1.4 Veillonellaceae

The representative family of the Negativicutes is very similar to the Acidaminococcacae, as they are found predominantly in the digestive tracts of humans (Whitman, 2011) and have been implicated in disease states on rare occurrences (Bhatti and Frank, 2000). Vellonellaceae includes six known genera: The type genus of the group is *Veillonella*, with 14 different species being isolated and described from humans and rodents (Euzeby, 1997). The type species, *Veillonella parvula*, is found in the gut and oral cavity of humans and plays a predominant role as a secondary colonizer of the teeth and mouth (Griffen *et al.*, 2012) by forming a mutualistic biofilm with *Streptococcus mutans* and other species (Periasamy and Kolenbrander, 2010). Interestingly *Veillonella* may play a role in childhood immune development as epidemiological studies have demonstrated a negative correlation with *Veillonella* and asthma (Arrieta *et al.*, 2015), bronchiolitis (Hasegawa *et al.*, 2016), and autism (Strati *et al.*, 2017).

4.2 Halanaerobiales

The second clade of diderm Firmicutes is the Halanaerobiales, a more recently discovered group of Bacteria (Mavromatis *et al.*, 2009) (Figure I.26). All members of the group are strict anaerobes and moderate halophiles. Most members of the order were isolated from saline lake sediments with some species being isolated from alkaline and acidic environments (Whitman, 2011). As most halophiles are in fact not Bacteria but Archaea, these bacteria have been studied for their adaptation to these environments. Very little is known about the cell structure in this group, however, they do appear to have a diderm cell structure similar to the Negativicutes.

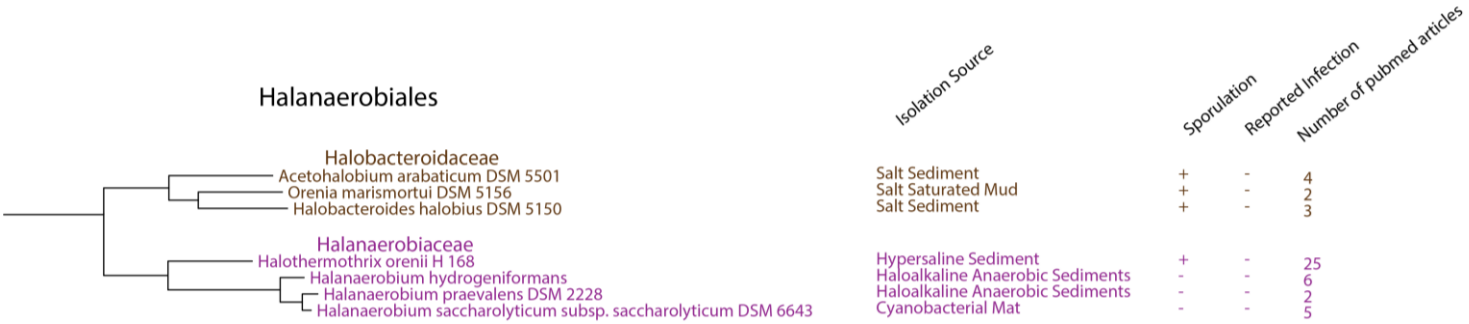

Figure I.26. Halanaerobiale Genomes Available as of October 2014.
Two families have been assigned (unique colours) and the strain details and number of PubMed articles are listed. Phylogeny was taken from (Antunes *et al.*, 2016)

What little work that has been done on this clade is incredibly interesting and has great potential, however, the studied information is very specific to a single topic that in turn are from many areas. One of the more interesting features is that *Acetohalobium arabaticum* encodes a twenty first-amino acid pyrrolysine, which is produced in the presence of trimethylamine (Prat *et al.*, 2012). Another interesting feature is that *Orenia* species have the capabilities to reduce various metals (Dong *et al.*, 2016). Halanaerobiales have also been found to cause corrosion in fracking operations by souring the gas (Liang *et al.*, 2016). Aside from these interesting cases, most research has focused on *Halothermothrix* and utilizing their unique enzymes for their temperature and salt resistant capabilities. For a review please see (Bhattacharya and Pletschke, 2014) or (Roush *et al.*, 2014).

4.3 Envelope Characteristics of Diderm Firmicutes

Most of the diderm Firmicutes test negative for the Gram stain and positive for the similar KOH test (Whitman, 2011). This test is not as reliable as once thought, though as many of the monoderm Firmicutes test the same as the diderm Firmicutes (Whitman, 2011). The classic Bladen (Bladen and Mergenhagen, 1964) study previously described was the first thin section microscopy of the diderm Firmicutes, however, it was not the last. Both the Halanaerobiales and the Negativicutes have had thin section microscopy (Figure I.27). More recently an in depth study of the sporulation on the termite gut bacterium *Acetonema longum* utilized cryo-tomography to produce the highest resolution pictures of a Negativicute (Tocheva *et al.*, 2011).

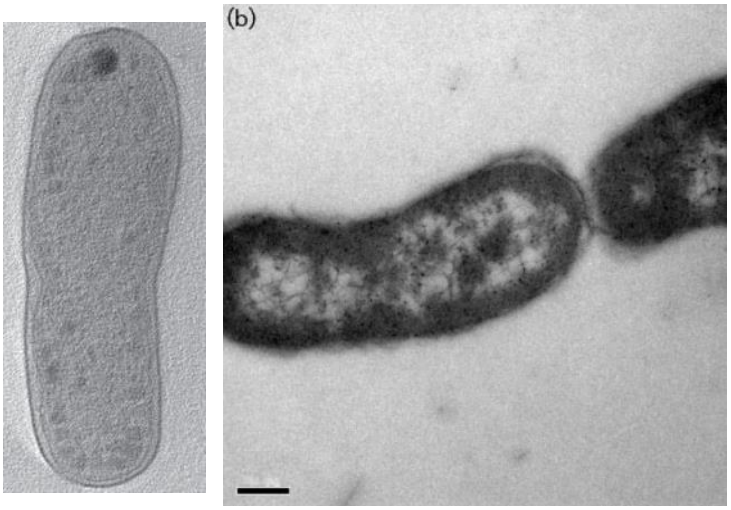

Figure I.27. Thin Section Microscopy of Diderm Firmicutes
Outer membrane envelope structure of the *Acetonema longum* (left) (Tocheva *et al.*, 2011) and *Halanaeroubium sehlinese* (right) (Abdeljabbar *et al.*, 2013)

One of the features of the diderm envelope that has been well studied is the mechanism of OM attachment. This method is completely different from that of *E. coli*. *E. coli* utilizes Braun's lipoprotein (Lpp) for attachment of the OM. In Selenomonas it was discovered that the peptidoglycan was modified with polyamines, such as putrescine, for attachment of the OM (Kojima et al., 2010). Later on, a protein, Mep45 or OmpM, was determined to bind the PG and integrate into the OM (Kojima et al., 2016). This topic is covered in depth in Results section 2.2.

The LPS of various Negativicutes has been characterized throughout the years, usually with a focus on the immunogenicity of the molecule. The LPS of *V. parvula* was recently found to activate the Toll-like receptor pathways and stimulated tumour necrosis factor alpha and interleukin-6 (Matera *et al.*, 2009). The LPS of some species has been characterized, with the best characterized belonging to the beer spoilage bacteria *Pectinatus* (Helander *et al.*, 2004). The LPS of *Pectinatus* species were found to have unique features including extensive heterogeneity within strains, a non-repeating polysaccharide that replaces the O-antigen, and an atypical lipid-A (Figure I.28). Other Negativicutes have had their LPS isolated including *A. longum* (Tocheva *et al.*, 2011) however no structure was determined. Halanaerobiale LPS has only recently had its LPS characterized. It was found that the deep sea *Halobacteroides lacunaris* possessed rough type LPS that inhibited the toxicity of *E. coli* LPS (Lorenzo *et al.*, 2017).

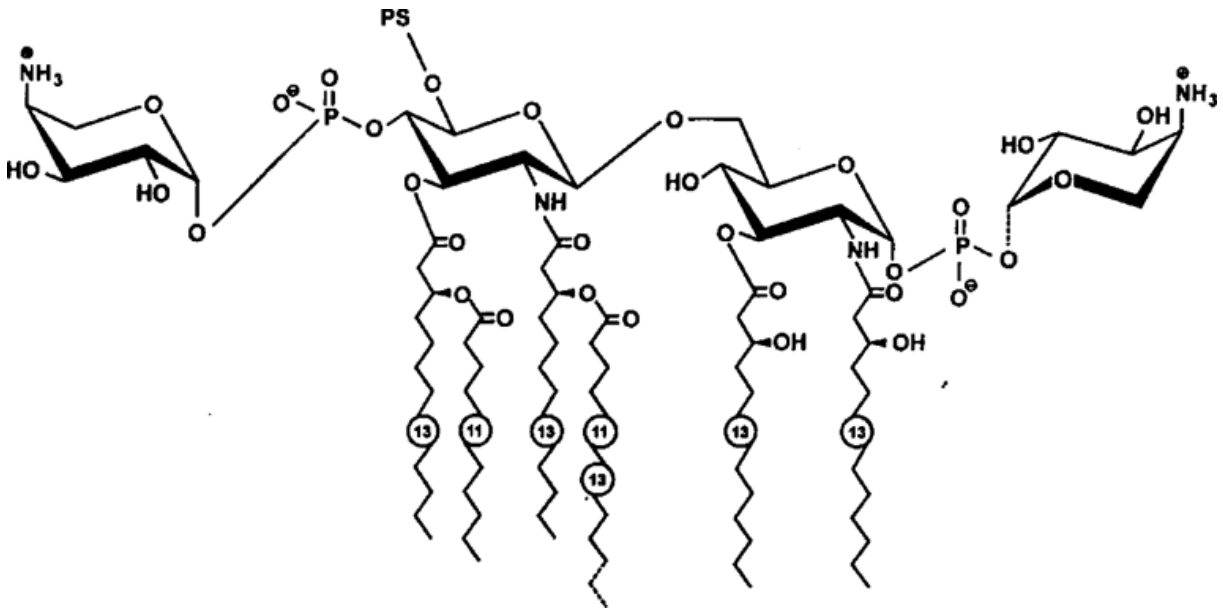

Figure I.28. Negativicute LPS structure
Structure of the lipid A component of S-form LPS of *P. cerevisiiphilus* and *P. frisingensis* (Helander *et al.*, 2004)

The most exciting and important aspect of this research is the close ties that the Negativicutes have to sporulation. This not only confirms their position as Firmicutes, it also has close ties to OM biogenesis.

4.4 Sporulation

Spore formation is a defining and conserved process among the Firmicutes for protection and stasis during unfavourable environmental conditions (Figure I.28).

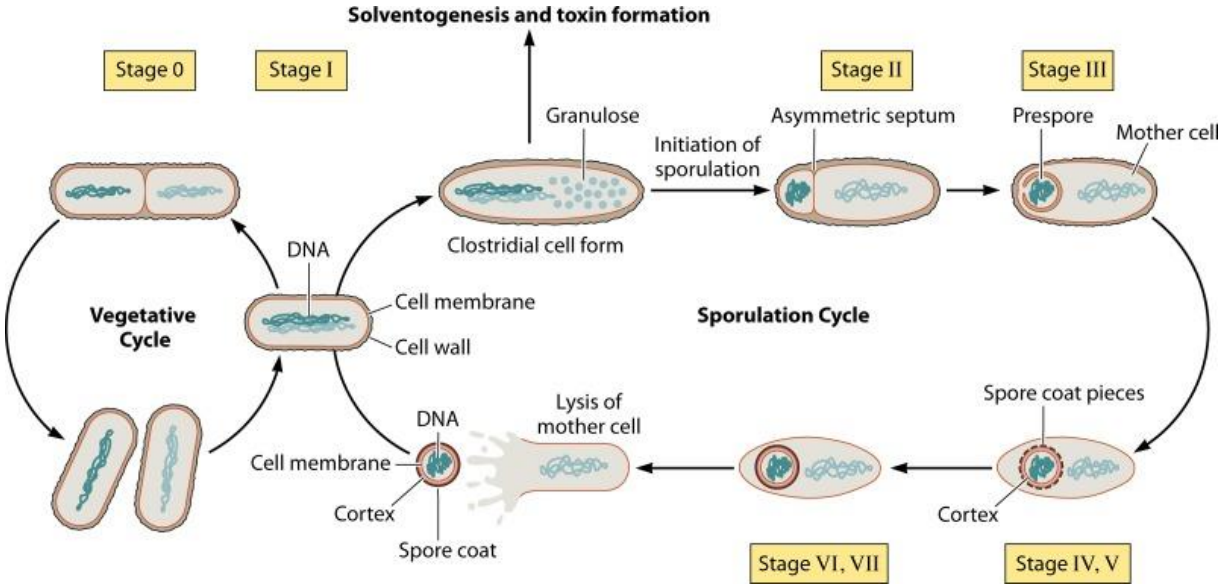

Figure I.29. Sporulation Cycle of Firmicutes
Sporulation starts when a vegetative cell (Stage 0) experiences unfavourable conditions sporulation genes are triggered (Stage I). The cell undergoes asymmetric division (Stage II) and a prespore is formed within the mother cell (Stage III). The prespore is coated and cellular contents are degraded (Stage IV, V). The prespore develops and then emerges from the mother cell (Stage VI, VII). Taken with permission from (Al-Hinai *et al.*, 2015)

4.4.1 Overview of Sporulation

Within all sporulating Firmicutes sporulation can be observed to go through seven stages (stage I–VII) (Figure I.29). Stage 0 is the growth of vegetative cells done by all bacteria before the beginning of sporulation. Sporulation is initiated by the induction of the master regulator Spo0A through a single cascade. In stages I and II, the cell undergoes asymmetric cell division and forms two compartments, the smaller prespore compartment and the larger mother cell compartment. At the end of stage II, the forespore becomes similar to an intracellular diderm bacterium. This diderm bacteria is formed by two different enzymes which divide the PG and drag the membrane along to produce a double membraned bacterium. These enzymes are SpoIID and SpoIIP. They thin the septal PG starting from the centre and move towards the edges until a diderm daughter cell is formed (Gutierrez *et al.*, 2010; Takeshita *et al.*, 2014)

After this, stage IV has the PG layer thickened around the forespore. The spore coat, a complex protein structure formed on the forespore, is formed in stage V. In stage VI the spore becomes resistant to heat and UV radiation. The final stage of spore formation, Stage VII is release of the spore from the mother cell and exposure to the environment (Hilbert and Piggot, 2004).

4.4.2 Sporulation in Diderm Firmicutes

Recently the sporulation of a Negativicute, *Acetonema longum*, was investigated (Tocheva et al., 2011). It was found that the process of OM biogenesis in some Negativicutes is closely tied to sporulation. During sporulation, the bacteria follow the same process as monoderm Firmicutes (Figure I.30), yet retains the membrane outside of the PG cortex. This membrane starts as a CM and ends as the new OM. During the course of their excellent microscopy work it was found that the cortex contains thick PG, however, this PG is thinned to the width of traditional diderm bacteria during germination.

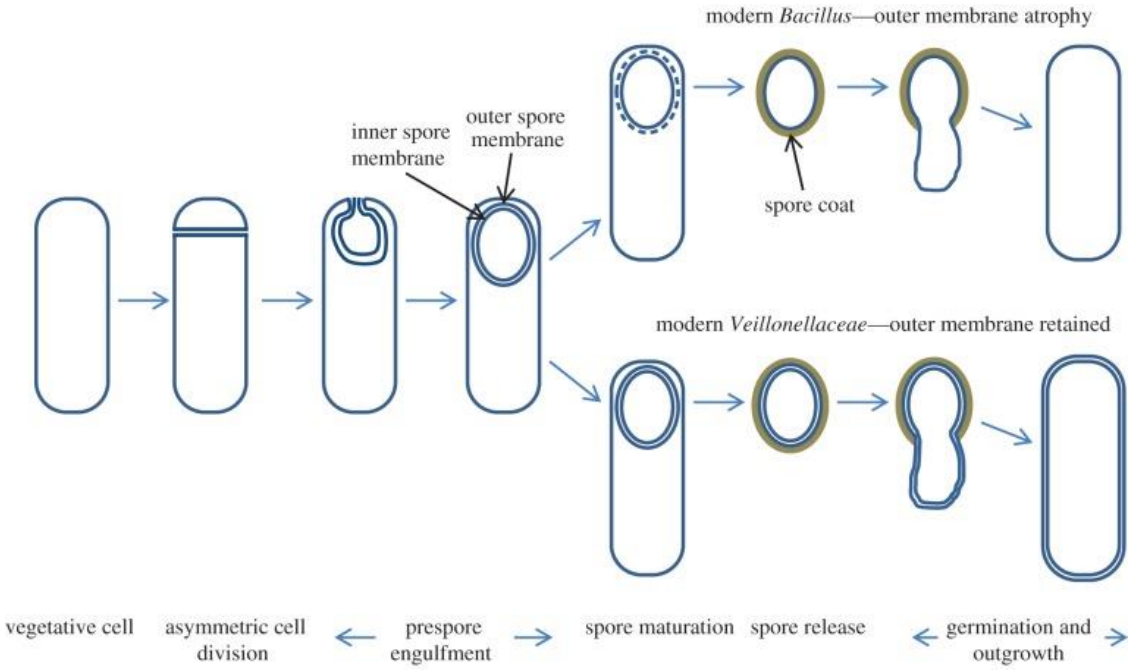

Figure I.30. Comparison of Sporulation of Diderm and Monoderm Firmicutes
Sporulation in monoderm Firmicute *B. subtilis* and a diderm Firmicute *Acetonema longum*. Taken with permission from (Errington *et al.*, 2016)

4.5 Evolutionary Relationships of Diderm Firmicutes

The great ant biologist E.O. Wilson once said that for every scientific question there exists an organism or group of organisms that are perfect for answering the question and for each organism a perfect question exists (Wilson, 2013). In our case, we have not only one group of atypically placed diderms, but two! Should we be able to resolve the phylogeny of the Firmicutes and the Firmicute OM, we will be able to determine if an OM is transferred or is an ancestral characteristic that is lost in monoderm clades. As this is an exciting question, I will present the past work on it.

4.5.1 Another Evolutionary Hypothesis: Sporulation in Diderm Firmicutes

The authors of the excellent sporulation paper of *A. longum* published an opinion paper (Tocheva *et al.*, 2016) shortly after to present their hypothesis on the evolutionary origin of the OM in Bacteria (Figure I.31). In their original paper (Tocheva *et al.*, 2011) they presented a tree of the Omp85 family of proteins. This tree was unresolved, however, it did not suggest horizontal transfer from any major Bacterial phylum. The authors propose that an ancestral organism was a sporulating monoderm. This organism developed the capability to keep the OM during sporulating similar to *A. longum.* This sporulating diderm is the LCA of all Bacteria. Then the LCA diversified with loss of the OM and sporulation in Actinobacteria and Chloroflexi, loss of the OM in monoderm Firmicutes, retention of both in some diderm Firmicutes and loss of sporulation in all the remaining bacteria. This origin story is plausible; it suggests that all Bacteria were once capable of sporulating yet no evidence exists for Firmicute-like sporulation outside of the Firmicutes phylum.

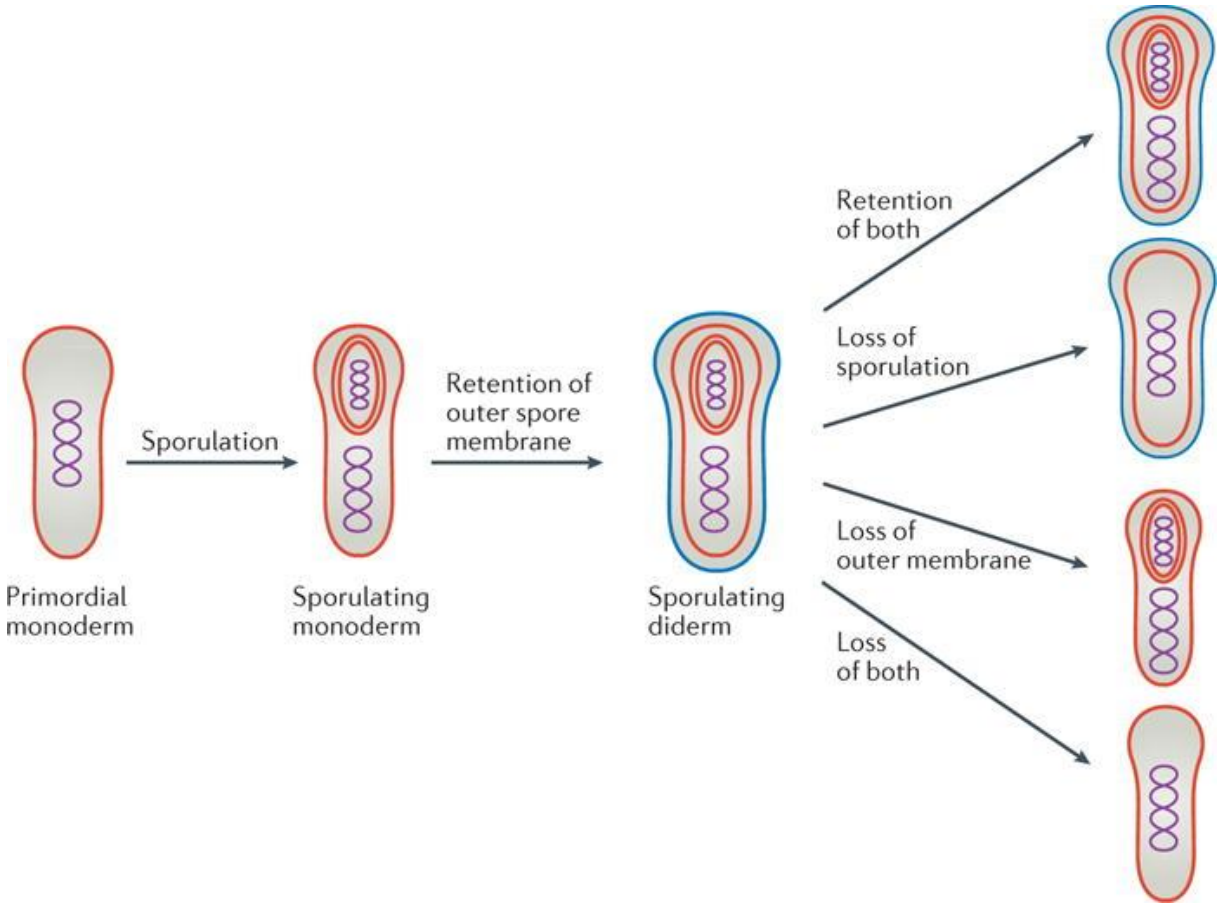

Figure I.31. The Tocheva Hypothesis
First, a monoderm ancestor developed sporulation. Then an OM was acquired through sporulation. This sporulating diderm is the LCA of Bacteria and developed into all modern lineages through loss of the OM, sporulation, or both. Taken with permission from (Tocheva *et al.*, 2011)

4.5.2 Transfer of Outer Membrane

In opposition to Tocheva et al. work demonstrating that the OM was not transferred from another clade, Campbell et al present the opposite hypothesis (Campbell *et al.*, 2014). In this work, they query each protein from the entire proteome of *A. intestini* via BLAST and record the top hits. Interestingly, they find that 7% of the proteome has a top hit within the Proteobacteria and the predominant portion of these genes relate to the function of the OM, suggesting that the OM can be transferred. This single genome data is interesting, but cannot be considered conclusive without further evidence. No comparable analysis has been done with the Halanaerobiales to see if a similar result is possible.

Another possibility is that the OM was transferred from Negativicutes to the Halanaerobiales or vice versa. This idea was presented based on a tree of LPS biosynthetic genes. In (Mavromatis *et al.*, 2009) they found that the Negativicutes and the Halanaerobiales grouped together and claimed this was due to a transfer. This data was not strongly supported due to the tree being constructed with neighbour joining methods and the inclusion of only one Negativicute and one Halanaerobe.

4.5.3 Key Questions on Firmicutes

Although the Firmicutes present the perfect opportunity for the study of the acquisition/loss of an OM, very little is known about the diderm Firmicute envelope. As I have stated, some work has been done on the Negativicutes cell envelope and structure, however very little is known about the Halanerobiales. Furthermore, it is not precisely known as to where these groups fall within the Firmicutes and what their relationship is with each other. I have endeavoured to answer these questions among others.

OBJECTIVES AND APPROACHES

The three goals of my Doctoral work are:

1. Elucidate the evolutionary history Firmicute envelopes
2. Describe the outer membranes of diderm Firmicutes
3. Find a mechanism of transition between diderm and monoderm Firmicutes.

1. Elucidate the evolutionary history of Firmicute envelopes

The evolutionary history of the Firmicute cell envelope will be studied through phylogeny. First, a robust phylogeny of the Firmicutes will be constructed that contains a diverse representation of both diderm and monoderm Firmicutes. This phylogeny will be constructed using a concatenation of ribosomal proteins and RNA polymerase. Then a phylogeny of the outer membrane will be constructed for the diderm Firmicutes. In this case, the biosynthetic genes of LPS will be concatenated. These trees will be compared for transfers both horizontal and vertical.

2. Describe the outer membrane of diderm Firmicutes

To achieve greater insight into the unique cell envelope architecture of diderm Firmicutes more needs to be done to see what the envelope possess and how this relates to known systems such as *E. coli*. Comparative genomics and thorough queries for gene families will enable us to see if they have the OM systems that have been well characterized before. In a biochemical fashion, the OM of *V. parvula* will be extracted and the proteins identified to verify the bioinformatics approach and identify new targets for study.

3. Find a mechanism of transition between diderm and monoderm Firmicutes.

We have hypothesized that the loss or gain of an OM gene could be due to a change in regulation of the PG. Therefore, I shall perform comparative genomics on all enzymes active in PG remodelling and synthesis to see if there is an enzyme that is present in all monoderm bacteria and absent from diderm or vice versa. Should this be successful I will attempt to express/delete the gene in *V. parvula* in hopes of replicating this transition.

RESULTS

1 Elucidate the evolutionary history of Firmicutes envelopes.

The first goal of my thesis was to understand the evolutionary history of the cell envelope. This is what I initially started my doctorate work with and it is how I have ended it. As such here are two articles that will tell the story of cell envelope history in the Firmicutes.

1.1 Phylogeny of Negativicutes, Halanaerobiales

The first work of my doctorate was started by a postdoctoral scientist, Dr Antunes and we worked on much of this article together and discovered the history of the Negativicutes, Halanaerobiales, and monoderm Firmicutes together. Dr Antunes had begun this work by constructing a local databank of 212 Firmicutes representing all major clades and groups. She proceeded to construct a ribosomal phylogeny that proved better than any available in the literature.

Together we proceeded on the annotation of the genes. She initially proceeded with an automatic HMM-based approach to search for all major diderm genes. I was able to enhance these automated results through the use of syntany and conserved domains to provide an annotation of diderm genes within the Firmicutes that has never been done to this level of completeness in this many genomes. In addition, this description enabled us to declare the absence of diderm genes within the monoderm lineages.

We then looked at individual systems to use for construction of an OM phylogeny. We ideally wanted a system that contained both monoderm and diderm genes in synteny. I constructed the syntany for the Flagella and T2SS (See Appendix 1), however, she could not construct a robust phylogeny with these genes. She attempted to construct the phylogeny with BAM genes, yet they proved to be very uninformative on the sequence level. After these trials, we moved onto the diderm cluster which we annotated and used the LPS genes to construct a phylogeny. Our first phylogeny of LPS concatenated *lpxABCDK*, *waaA*, and *msbA*, however, we noticed that we could improve resolution with just the *lpxABCD* genes. The LPS genes were queried by Dr Brochier and I constructed the final tree.

This data was complemented with the beautiful microscopy of Dr Klingl. Luisa had grown a Negativicute and one Halanaerobiale and they were imaged using the same technique. This clearly shows that both of these groups have a similar envelope architecture and firmly establishes them as diderm.

The writing process was highly involved and took several iterations. Many of the analysis that were performed did not make it into the final version. Upon submission and receiving the reviewers' comments I performed hypothesis testing and congruency tests to discount any possibility that the Negativicutes and the Halanaerobiales form a monophyletic

group, or that the proteins should not be used in the concatenation. This work was done in partnership with Dr.Criscuolo.

The takeaway messages of this paper is that the diderm envelope is an ancestral characteristic of the Firmicutes and the OM was lost multiple times independently. This work was published in eLIFE and was the focus of a companion article (Muñoz-Gómez and Roger, 2016), was featured on the ASM podcast TWM, and multiple blogs.

eLIFE elifesciences.org

RESEARCH ARTICLE

Phylogenomic analysis supports the ancestral presence of LPS-outer membranes in the Firmicutes

Luisa CS Antunes[1†], Daniel Poppleton[1†], Andreas Klingl[2], Alexis Criscuolo[3], Bruno Dupuy[4], Céline Brochier-Armanet[5], Christophe Beloin[6], Simonetta Gribaldo[1*]

[1]Unité de Biologie Moléculaire du Gène chez les Extrêmophiles, Département de Microbiologie, Institut Pasteur, Paris, France; [2]Plant Development and Electron Microscopy, Department of Biology I, Biocenter LMU, Munich, Germany; [3]Bioinformatics and Biostatistics Hub, Institut Pasteur, Paris, France; [4]Laboratoire Pathogenèse des Bactéries Anaérobies, Département de Microbiologie, Institut Pasteur, Paris; [5]Laboratoire de Biométrie et Biologie Évolutive, University Lyon I, Villeurbanne, France; [6]Unité de Génétique des Biofilms, Département de Microbiologie, Institut Pasteur, Paris, France

Abstract One of the major unanswered questions in evolutionary biology is when and how the transition between diderm (two membranes) and monoderm (one membrane) cell envelopes occurred in Bacteria. The Negativicutes and the Halanaerobiales belong to the classically monoderm Firmicutes, but possess outer membranes with lipopolysaccharide (LPS-OM). Here, we show that they form two phylogenetically distinct lineages, each close to different monoderm relatives. In contrast, their core LPS biosynthesis enzymes were inherited vertically, as in the majority of bacterial phyla. Finally, annotation of key OM systems in the Halanaerobiales and the Negativicutes shows a puzzling combination of monoderm and diderm features. Together, these results support the hypothesis that the LPS-OMs of Negativicutes and Halanaerobiales are remnants of an ancient diderm cell envelope that was present in the ancestor of the Firmicutes, and that the monoderm phenotype in this phylum is a derived character that arose multiple times independently through OM loss.

DOI: 10.7554/eLife.14589.001

***For correspondence:** simonetta.gribaldo@pasteur.fr

[†]These authors contributed equally to this work

Competing interests: The authors declare that no competing interests exist.

Funding: See page 17

Received: 22 January 2016
Accepted: 21 July 2016
Published: 31 August 2016

Reviewing editor: Jonathan Eisen, UC Davis, United States

Introduction

The bacterial envelope is one of the oldest and most essential cellular components, involved in key housekeeping functions such as physical integrity, cell division, motility, substrate uptake and secretion, and cell-cell communication (***Silhavy et al., 2010***). Yet, bacteria show substantial differences in their cell envelope architectures, among which the most dramatic one is the presence of one (monoderm) or two (diderm) membranes (***Sutcliffe, 2010***). The study of cell envelope architecture has been mostly narrowed to the Firmicutes and the Gammaproteobacteria as textbook examples of monoderm and diderm bacteria, respectively. In *Bacillus subtilis*, teichoic and lipoteichoic acids are embedded in a thick peptidoglycan wall, while in *Escherichia coli* a thin peptidoglycan layer is surrounded by an outer membrane (OM) whose biogenesis and functioning involve a complex system of synthesis and transport for LPS, lipoproteins, and OM proteins (OMPs) (***Silhavy et al., 2010***).

The transition between monoderm and diderm cell envelopes must have been a significant and complex process in the evolutionary history of Bacteria. Two major hypotheses have been largely discussed in the literature, which can be generally defined as *diderm-first* (***Cavalier-Smith, 2006***)

eLife digest The cell envelope is one of the evolutionarily oldest parts of a bacterium. This structure – made up of a cell wall and either one or two cell membranes – surrounds the bacterial cell, maintaining the cell's structure and providing an interface through which bacteria can sense their environment and communicate.

Bacteria can be broadly classed based on the number of cell membranes that their envelope consists of. Bacteria that have a single cell membrane are known as "monoderm", whereas those with two membranes are termed "diderm". The number of membranes that bacteria have can affect how well they resist antibacterial compounds. When, how and why bacteria switched between monoderm and diderm cell envelopes are some of the major unanswered questions in evolutionary biology.

The textbook example of a monoderm cell envelope can be found in bacteria called Firmicutes. This group includes some notoriously harmful bacteria such as *Staphylococcus*, which can cause conditions ranging from abscesses to pneumonia. However, some Firmicutes possess two cell membranes. It was unclear how these unusual diderm Firmicutes developed a second membrane, and how they are related to their monoderm relatives.

Antunes, Poppleton et al. set out to answer these questions by analyzing the information contained in the thousands of bacterial genomes that have already been described. The results indicate that Firmicutes originally had diderm envelopes, and that species with monoderm envelopes arose independently several times through the loss of their outermost membrane.

Future work is needed to investigate the driving forces and the precise mechanism that led most Firmicutes to lose their outer membrane. Also, further characterization of diderm Firmicutes will provide key information about the biology of these poorly understood bacteria.

DOI: 10.7554/eLife.14589.002

and *monoderm-first* (**Gupta, 2011**; **Lake, 2009**) scenarios. The fact that the majority of phyla seem to possess two membranes might favor the *diderm-first* scenario, although the actual diversity of cell envelopes in Bacteria remains largely unexplored (**Sutcliffe, 2010**). However, the lack of a robustly resolved phylogeny for Bacteria, notably the uncertainty on its root and the nature of the earliest branches, has left the relationships between diderm and monoderm phyla unclear, and not allowed to define in which direction and how many times this transition occurred.

In this respect, the Negativicutes (**Marchandin et al., 2010**) represent an interesting case: while belonging phylogenetically to the classical monoderm Firmicutes, they surprisingly display a diderm cell envelope with an OM and LPS (**Delwiche et al., 1985**; **Vos et al., 2009**). The Negativicutes have been identified in various anaerobic environments, such as soil and lake sediments, industrial waste, and animal digestive tract (**Vos et al., 2009**). Their best-characterized member is *Veillonella*, first described in 1898 by Veillon and Zuber (**Veillon and Zuber, 1898**). Curiously, the very first observation and use of the term 'outer membrane' has been based on studies of *Veillonella* (**Bladen and Mergenhagen, 1964**). *Veillonella* is one of the most abundant components of the human oral flora (**Tanner et al., 2011**), and a common inhabitant of the intestinal microbiome (**van den Bogert et al., 2013**). Together with other gut microbes, it has been recently associated with maturation of the immune system and partial protection of asthma in infants (**Arrieta et al., 2015**), but can also develop into an opportunistic pathogen (**Hirai et al., 2016**). Several other Negativicutes members such as *Dialister, Selenomonas, Mitsuokella,* and *Anaeroglobus* show increased incidence in oral tract disease linked to biofilm formation (**Griffen et al., 2012**) and involvement in other infections (**Wang et al., 2015**). Very little experimental data is available on the nature of the diderm cell envelope of Negativicutes. In *Selenomonas ruminantium* the abundant OmpM protein appears to replace the important function of Braun's lipoprotein in anchoring the OM to the cell peptidoglycan through a link with cadaverine (**Kojima et al., 2010**).

How the OM originated in the Negativicutes represents an evolutionary conundrum. Recently, Tocheva and colleagues analyzed the sporulation process in the Negativicute *Acetonema longum* by cryoelectron microscopy (**Tocheva et al., 2011**). They showed that, while an outer membrane forms only transiently during sporulation in classically monoderm Firmicutes such as *Bacillus subtilis*, it is

retained in *A. longum* leading to its diderm phenotype (**Tocheva et al., 2011**). This study provided the first experimental support for the hypothesis that the bacterial OM could have initially evolved in an ancient sporulating monoderm bacterium (**Dawes et al., 1980**; **Errington, 2013**; **Vollmer, 2012**). Moreover, a phylogenetic tree of the essential Omp85 protein family for proteins insertion in the outer membrane, although largely unresolved, did not show the Negativicutes as emerging from any specific diderm phylum (**Tocheva et al., 2011**). The authors speculated that the OM of Negativicutes was not acquired by horizontal gene transfer but was already present in the ancestor of Firmicutes and would have been lost in the other members of this phylum, although it remained unclear when and how many times this would have occurred (**Tocheva et al., 2011**). In contrast, a recent analysis of the genome of the Negativicute *Acidaminococcus intestini* revealed that as much as 7% of the BLAST top hits were from Proteobacteria, the majority of which corresponded to functions related to OM biogenesis, concluding to a possible acquisition of the OM in Negativicutes by horizontal gene transfer (**Campbell et al., 2014**).

Interestingly, the Negativicutes are not the only diderm lineage in the Firmicutes. The Halanaerobiales are a poorly studied group of moderate halophilic, strictly anaerobic Firmicutes that were isolated from saline environments such as lake and lagoon sediments, and oil reservoirs (**Oren, 2006**; **Roush et al., 2014**). Similarly to the Negativicutes, they display a diderm-type cell envelope, with a thin peptidoglycan and an outer membrane (**Cayol et al., 1994**; **Zeikus et al., 1983**; **Zhilina et al., 1992**, **Zhilina et al., 2012**). When analyzing the first sequenced genome of a member of Halanaerobiales, *Halothermothrix orenii*, Mavromatis and colleagues identified a number of OM markers, suggesting the presence of an LPS-diderm cell envelope homologous to the one of Negativicutes and other diderm bacteria (**Mavromatis et al., 2009**). In contrast to the few analyses on Negativicutes, no experimental data are available on the characteristics of the OM in the Halanaerobiales.

The existence of two diderm lineages in the Firmicutes provides a fantastic opportunity to clarify the monoderm/diderm transition in this major bacterial phylum. However, the origins and evolutionary relationships between the OM of Halanaerobiales and Negativicutes have been unclear. In fact, no Halanaerobiales were present in the analysis of Tocheva (**Tocheva et al., 2011**). Mavromatis et al. built a tree from the combined analysis of the genes coding for LPS, which showed a clustering of Halanaerobiales and Negativicutes, leading the authors to propose a horizontal gene transfer of the OM between these two lineages (**Mavromatis et al., 2009**). However, the sequenced genome of only one member of Halanaerobiales and one of Negativicutes were available at the time, and the LPS tree was largely unresolved (**Mavromatis et al., 2009**). Moreover, current phylogenies of the Firmicutes have been unclear with respect to the relationships between Negativicutes and Halanaerobiales. The Negativicutes have been alternatively indicated as branching within Clostridia (**Yutin and Galperin, 2013**; **Mavromatis et al., 2009**; **Vesth et al., 2013**) or at the base of Bacilli (**Kunisawa, 2015**). The phylogenetic placement of Halanaerobiales remains also uncertain, as they have been assigned either to Class Clostridia (**Cayol et al., 1994**), as a deep branch in the Firmicutes (**Mavromatis et al., 2009**; **Vos et al., 2009**; **Kunisawa, 2015**), or left unresolved (**Yutin and Galperin, 2013**). Finally, no detailed genomic analysis has been carried out to infer and compare the characteristics of the cell envelopes of several Halanaerobiales and Negativicutes.

The large number of Negativicutes and Halanaerobiales genomes currently available prompted us to carry out a global phylogenomic study. This allowed to robustly clarifying the relative placement of Negativicutes and Halanaerobiales within the Firmicutes, to assess the evolutionary relationships of their cell envelopes, and to perform in depth comparative analysis to understand the characteristics of key OM-related processes in these two lineages. Our results provide robust support for an emergence of monoderm Firmicutes from diderm ancestors via multiple independent losses of the OM.

Results

Electron microscopy of the diderm cell envelopes of Halanaerobiales and Negativicutes

Although the presence of an OM has been previously shown by electron microscopy for members of Negativicutes (e.g. **Tocheva et al., 2011**) and Halanaerobiales (e.g. **Zhilina et al., 2012**), these images have been obtained separately and with different techniques, making difficult their

comparison. We therefore obtained electron microscopy images of one representative of Negativicutes (*Megamonas rupellensis*) and one of Halanaerobiales (*Halanaerobium saccharolyticum*). We used transmission electron microscopy (TEM) following high-pressure freezing, freeze substitution, plastic embedding and ultrathin sectioning of the samples (see Materials and Methods). The application of high-pressure freezing in combination with appropriate freeze-substitution protocols facilitates the ultrastructural analysis of microorganisms and their membranes and also results in densely and homogeneously packed cytoplasm (**McDonald, 2007**; **McDonald et al., 2007a**; **Rachel et al., 2010**).

Ultrathin sections of high-pressure frozen cells of *M. rupellensis* and *H. saccharolyticum* confirmed the presence of clearly diderm-type cell envelope architecture in both strains (**Figure 1**). In a cross section from inside to outside, the densely packed cytoplasm of *M. rupellensis* is surrounded by a cytoplasmic membrane followed by the periplasm with a thin peptidoglycan layer and an OM (**Figure 1A**). Furthermore, pilus-like structures could be detected, as well as another electron dense layer outside the OM, which might correspond either to the lipopolysaccharide (LPS) or an S-layer. Thin sections of *H. saccharolyticum* also revealed a diderm cell envelope with a densely packed cytoplasm enclosed by a membrane surrounded by a relatively electron lucent periplasm so that the thin line representing the peptidoglycan is clearly visible (**Figure 1B**). For both organisms, in some cases the periplasm appeared inflated (**Figure 1C and D**), which was most likely caused by a preparation artifact due to swelling of the cells in the freeze substitution process. This effect nevertheless enabled us to observe the peptidoglycan much better as compared to cells without that artifact.

Robust phylum-level phylogeny of the Firmicutes supports distinct origins of Halanaerobiales and Negativicutes

We gathered homologues of 47 ribosomal proteins from a local database of 205 Firmicutes taxa and 13 bacteria belonging to eight major phyla as outgroup (Materials and methods). We did not include the Tenericutes in the analysis, because their reduced genomes and fast evolutionary rates are likely to cause artifacts in deep phylogenies, but it is known that they phylogenetically belong to the Bacilli (**Davis et al., 2013**). We assembled the 47 ribosomal proteins into a large concatenated dataset (5551 amino acid characters) and carried out Bayesian analysis with a sophisticated site-heterogeneous model of protein evolution (CAT) that allows each site to evolve under its own substitution matrix and is robust against tree reconstruction artifacts that frequently affect deep phylogenies (**Lartillot and Philippe, 2004**). The Bayesian tree was well resolved at most nodes (Posterior Probabilities (PP) > 0.95, **Figure 2**). Despite the weak signal and the stochastic errors frequently associated to small proteins such as ribosomal ones, topology congruence tests on individual markers showed a largely congruent phylogenetic signal, especially at high rank taxonomy level, justifying their combined analysis (**Figure 2—figure supplement 1** and Materials and methods). Maximum Likelihood (ML) analysis of the same concatenated dataset and the site-homogeneous LG model (**Le and Gascuel, 2008**) gave a largely consistent topology although it was much less resolved, especially at deep nodes (**Figure 2—figure supplement 2**). With respect to previous analyses, the relative placement of Negativicutes and Halanaerobiales in the Firmicutes phylogeny was robustly resolved (**Figure 2**). In fact, the Negativicutes branched within Class Clostridia, specifically related to Peptococcaceae and other incertae-sedis clostridial families (PP = 1, **Figure 2**). This placement is consistent with previous analyses, although performed with less taxa (**Yutin and Galperin, 2013**; **Mavromatis et al., 2009**). As opposed to the diderm nature of Negativicutes, the members of Peptococcaceae have monoderm phenotype (**Vos et al., 2009**) and no homologues of OM markers.

In contrast, the Halanaerobiales emerged as a distinct, well-supported, and deep-branching lineage of the Firmicutes (PP = 0.99, **Figure 2**), robustly grouped with the order Natranaerobiales (PP = 1, **Figure 2**). This clustering was also observed in a previous analysis performed with only one member of Halanaerobiales and one of Natranaerobiales, and its position in the Firmicutes phylogeny was left unresolved (**Yutin and Galperin, 2013**). Natranaerobiales are a poorly known group of moderately halophilic Firmicutes that appear monoderm under the microscope (**Mesbah et al., 2007**) and have no homologues of OM markers.

In order to verify the robustness of the distinct branching of the two diderm Firmicutes lineages, we ran AU tests on 12 topologies alternative to the Bayesian ribosomal protein concatenate tree, where Negativicutes or Halanaerobiales were moved 'up and down' the six nodes separating them (N_1-N_6 for the topologies involving moving the Negativicutes; H_1-H_6 for the topologies

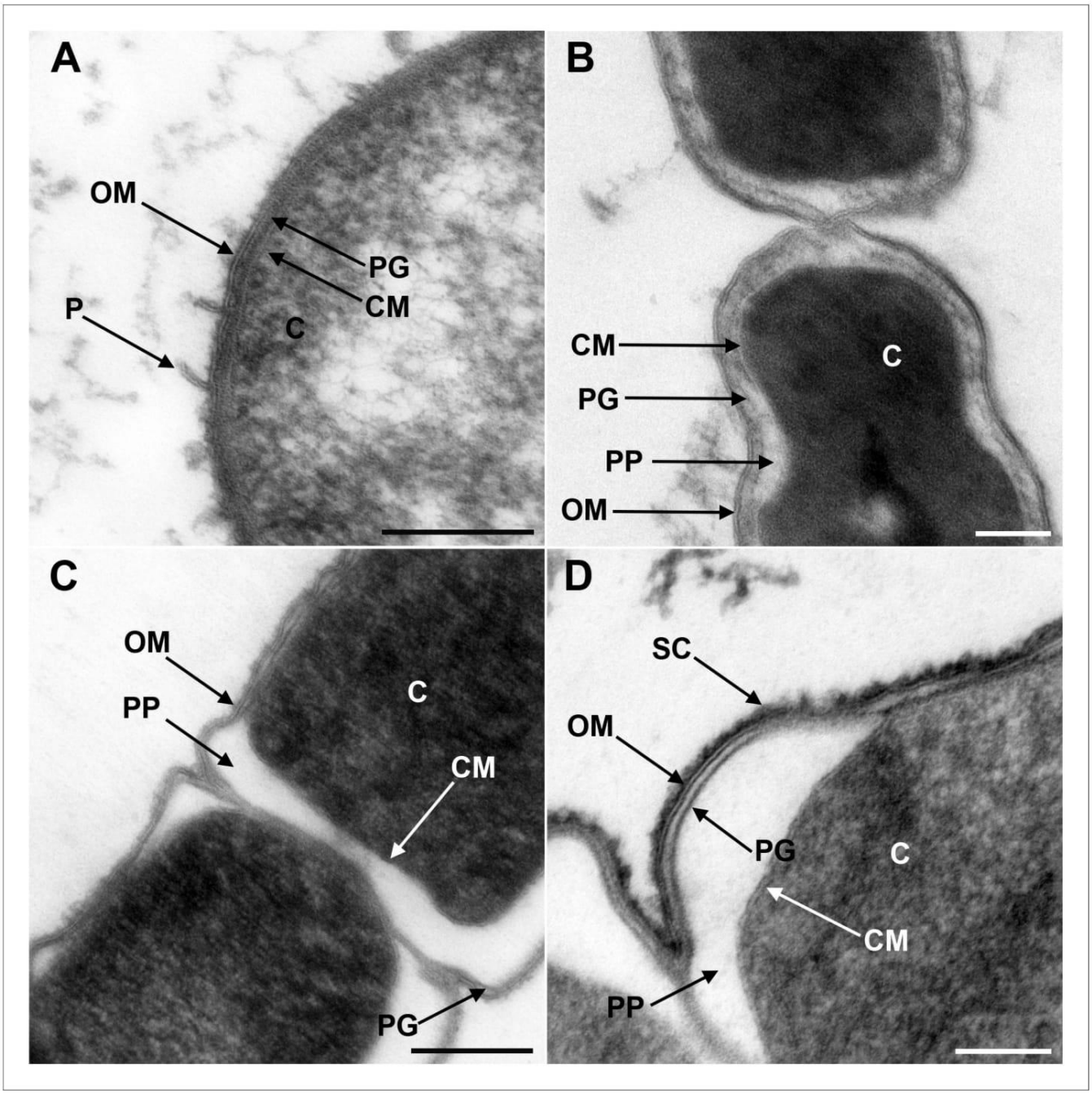

Figure 1. Transmission electron microscopy of a member of Negativicutes and a member of Halanaerobiales. Ultrathin sections of high-pressure frozen cells of the Negativicutes member *Megamonas rupellensis* (**A**,**C**), and the Halanerobiales member *Halanaerobium saccharolyticum* (**B**,**D**). A Gram-negative like cell wall architecture is visible for both taxa (**A**,**B**): a cytoplasmic membrane (CM) surrounding the cytoplasm (**C**), a thin peptidoglycan layer (PG), and an outer membrane (OM). Pili-like structures (P) are also visible in *M. rupelllensis*. In some cases and due to a preparation artifact caused by swelling of the cells, the OM detaches from the IM creating an enlarged periplasmic space (PP) between two dividing cells (**C**,**D**). In these cases, the peptidoglycan becomes more apparent as it is also the case for an electron dense surface coat (SC), which might represent lipopolysaccharide (LPS) or a potential S-layer. Scale bars: 200 nm (**A**,**C**) and 100 nm (**B**,**D**).

DOI: 10.7554/eLife.14589.003

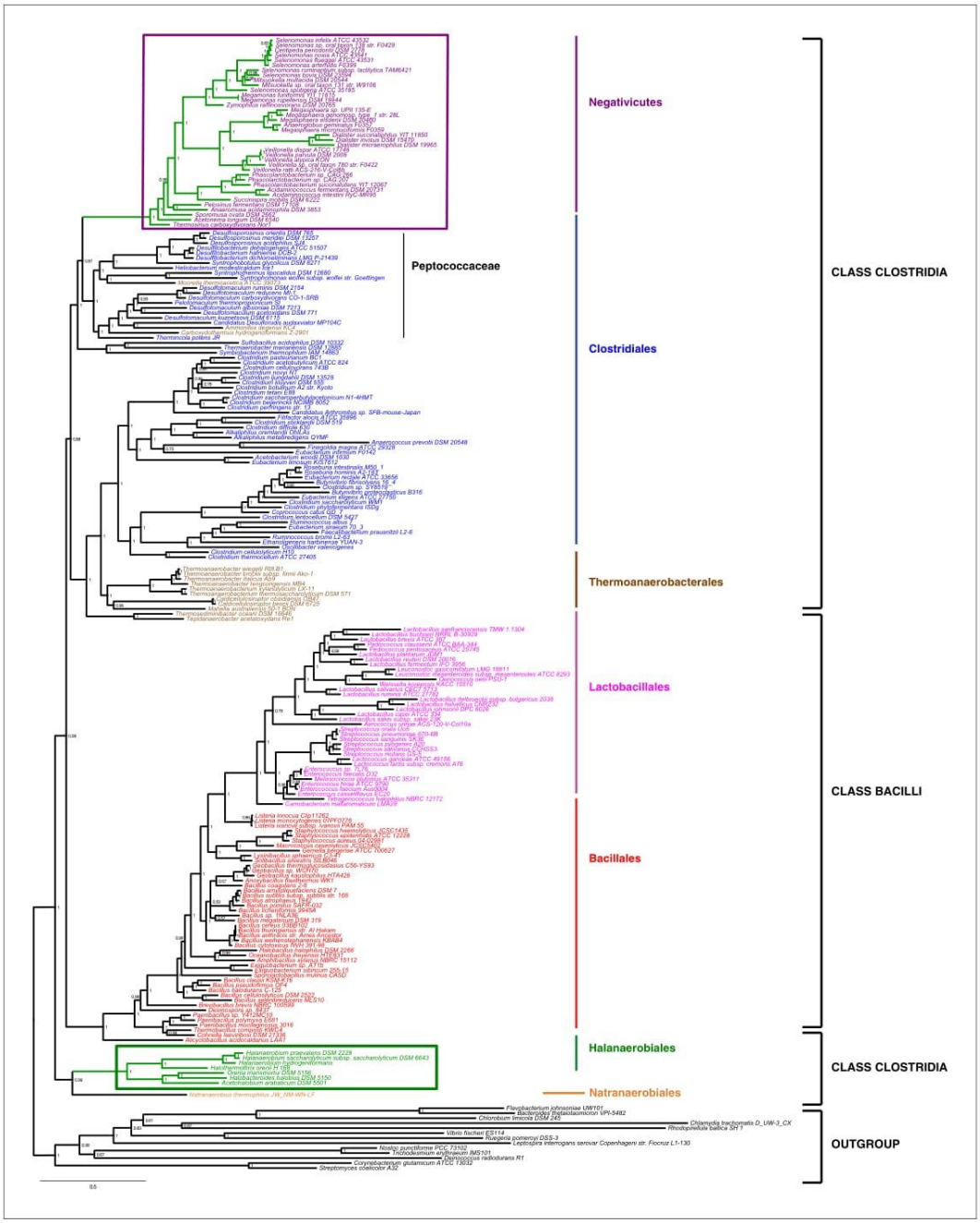

Figure 2. Phylum-level phylogeny of the Firmicutes. Bayesian phylogeny of the Firmicutes based on a concatenation of 47 orthologous ribosomal proteins comprising 5551 amino acid positions and the CAT+GTR+Γ4 model. Values at nodes represent Bayesian posterior probabilities. The scale bar represents the average number of substitutions per site. For details on analyses, see Materials and methods.

DOI: 10.7554/eLife.14589.004

The following figure supplements are available for figure 2:

Figure supplement 1. Results of IC congruence test for the 47 ribosomal proteins.

DOI: 10.7554/eLife.14589.005

Figure supplement 2. Maximum likelihood phylogeny of the Firmicutes.

DOI: 10.7554/eLife.14589.006

Figure supplement 3. Results of AU test for 12 alternative topologies.

DOI: 10.7554/eLife.14589.007

involving moving the Halanaerobiales ***Figure 2—figure supplement 3***, Materials and methods and Additional Data). Unfortunately, topology testing is currently only available in a Maximum Likelihood framework with site-homogeneous models. Accordingly, these tests should reflect the poor resolution of the deep nodes of the maximum likelihood tree (***Figure 2—figure supplement 2***). Nevertheless, the Bayesian topology of ***Figure 2***(H_0N_0) was the preferred one (***Figure 2—figure supplement 3***). Two alternative topologies only (N_1 and N_2) were not rejected by the data, where the Negativicutes branched earlier in the Clostridia. Importantly, the two alternative topologies presenting a clustering of Halanaerobiales and Negativicutes (H_6 and N_6) were strongly rejected by the data, as well as all topologies where the Halanaerobiales were moved away from the root of the Firmicutes tree (H_2-H_5) (***Figure 2—figure supplement 3*** and Additional Data), consistent with the separate origins of the two diderm lineages.

To sum up, our phylogenetic analysis shows that Halanaerobiales and Negativicutes have distinct evolutionary origins, and are each related to different monoderm Firmicutes lineages.

The LPS-OM of Negativicutes and Halanaerobiales are homologous structures with an ancient origin

In contrast to the distinct emergence of Halanaerobiales and Negativicutes in the Firmicutes, the presence in their genomes of markers related to OM biogenesis and functioning (***Campbell et al., 2014***; ***Mavromatis et al., 2009***; ***Tocheva et al., 2011***) clearly indicates that their diderm cell envelopes are homologous structures. However, as discussed in the Introduction section, the specific evolutionary relationships between the OMs of Halanaerobiales and Negativicutes have been unclear. Interestingly, synteny analyses revealed a large genomic locus that is conserved between Halanaerobiales and Negativicutes, and is not present in monoderm Firmicutes (***Figure 3*** and ***Supplementary file 1***). Other than LPS synthesis and transport (green), the genes belonging to this genomic locus encode a number of cell envelope systems, such as OMP assembly (blue), motility (light pink), OM-PG attachment (red), efflux (purple), but also a number of hypothetical proteins (brown), and proteins not known to be specifically related to the OM (white).

Such clustering is unusual as in *E. coli* for example these genes are scattered in different regions of the genome. However, the genes coding for the first four steps of LPS synthesis (*lpxABCD*) display a conserved synteny in diderm Bacteria at very large evolutionary distances, suggesting that they have similar evolutionary histories (***Opiyo et al., 2010***). Accordingly, synteny is also conserved

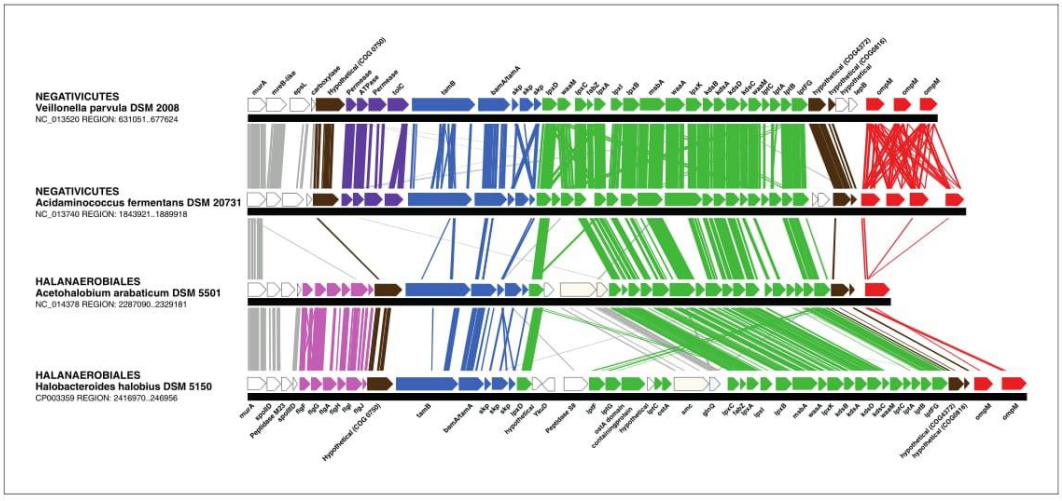

Figure 3. Conserved genomic locus for cell envelope components. Co-localization of the genes coding for LPS synthesis and transport, OMP assembly and structural OMPs in the Negativicutes and the Halanaerobiales. Representatives of the 2 families of Negativicutes and the 2 families of Halanaerobiales are shown (for full distribution and accession numbers see ***Supplementary file 1***). Genes are colored according to their functional class: LPS synthesis and transport (green), OMP assembly (blue), flagellum (light pink), OM-PG attachment (red), hypothetical (brown), efflux (purple) (see text for discussion). White boxes indicate proteins not known to being related to the OM or non-conserved proteins whose connection with the OM is unclear. The figure was obtained by EasyFig (***Sullivan et al., 2011***), where vertical lines represent BLAST hits with a cutoff of 0.0001.
DOI: 10.7554/eLife.14589.008

in Halanaerobiales and Negativicutes (***Figure 3*** and ***Supplementary file 1***). We therefore searched for these four core LPS genes (*lpxABCD*) in a local databank of 121 genomes representative of 30 major bacterial phyla (Materials and methods and ***Supplementary file 2***). As compared to what could be previously inferred from available genomic data (***Sutcliffe, 2010***; ***Opiyo et al., 2010***), we show the presence of homologues of the four core LPS coding genes in 26 major bacterial phyla, eight of which evidenced for the first time: Thermodesulfobacteria, Fibrobacteres, Ignavibacteria, Nitrospina, Chrysiogenetes, Cloacimonetes, Atribacteria, and Armatimonadetes (***Supplementary file 2***). This suggests that LPS-diderm cell envelopes might be even more widespread in Bacteria than currently thought, leaving only four major phyla that appear to lack the coding capacity for LPS: Thermotogae, Caldiserica, Chloroflexi/Thermomicrobia, and Actinobacteria (***Supplementary file 2***). We assembled the four core LPS protein homologues into a concatenated dataset (898 amino acid characters) also including Halanaerobiales and Negativicutes, and obtained a Bayesian tree with the CAT+GTR+Γ_4 evolutionary model (***Figure 4***, Materials and methods). In agreement with their conserved synteny, congruence tests showed that these four core LPS genes have a consistent phylogenetic signal at large evolutionary distances, in particular concerning the monophyly of major bacterial phyla, justifying their combined analysis (Materials and methods and ***Figure 4—figure supplement 1***). Consistently with the notorious difficulty in resolving the global phylogeny of Bacteria, the tree is not completely resolved. However, it is largely in agreement with bacterial systematics, showing the monophyly of major phyla (***Figure 4***). This pattern indicates that the core LPS genes were present in the ancestor of each of these diderm phyla, and that inter-phylum horizontal gene transfers were surprisingly rare during bacterial evolution. Consistently, the Halanaerobiales and Negativicutes also form a well-supported monophyletic cluster (PP = 1, ***Figure 4***), with internal branching pattern matching their respective reference species phylogeny shown in ***Figure 2***.

These results indicate that the LPS-OM of Halanaerobiales and Negativicutes do not have distinct origins, but rather that, similarly to the other main diderm bacterial phyla, they were inherited from their common ancestor, which is also the ancestor of all Firmicutes, in agreement with ***Tocheva et al., 2011***. The inclusion of a second diderm lineage in our analysis allows us to strengthen and extend this scenario, and to infer that present-day monoderm Firmicutes would have emerged from diderm ancestors via not less than five independent losses of the OM (***Figure 4B***). The two alternative topologies that were not rejected by AU tests do not affect the inference of a diderm ancestor and imply four and three independent OM losses, respectively (***Figure 2—figure supplement 3*** and Additional Data).

Outer membranes in a monoderm context

Our phylogenetic analyses suggest that the diderm cell envelopes of Halanaerobiales and Negativicutes might be the remnants of ancient bacterial structures that were inherited from the Firmicutes ancestor. In the absence of experimental characterization, exploration of genomic data can guide inferences on the nature of these atypical diderm cell envelopes. To this aim, we investigated a few key processes that are related to OM biogenesis and functioning and are shared between Negativicutes and Halanaerobiales. Because OM markers frequently display low sequence conservation or are part of large membrane-related protein families often preventing the building of robust phylogenies, we helped tentative annotation by merging information obtained from homology to known OM markers, the presence of specific protein domains, and genomic synteny. In this respect, the presence of the conserved OM locus helped annotation greatly and provided important insights into the unique nature of the cell envelopes of Negativicutes and Halanaerobiales, which show both specific characteristics as well as an intriguing combination of diderm and monoderm features (***Figure 5***).

Diderm Firmicutes synthesize and transport LPS to the OM

LPS is a complex glycolipid exclusively present in the outer leaflet of the OM. Although LPS can be very heterogeneous in bacteria, it has an overall conserved structure composed of a membrane-anchored hydrophobic domain (lipid A, or endotoxin), an oligosaccharide (inner and outer core), which can be extended with a distal polysaccharide (O-antigen) (***Wang et al., 2010***). The Lipid A-core portion (known as 'rough' LPS) and the O-antigen have independent pathways for their

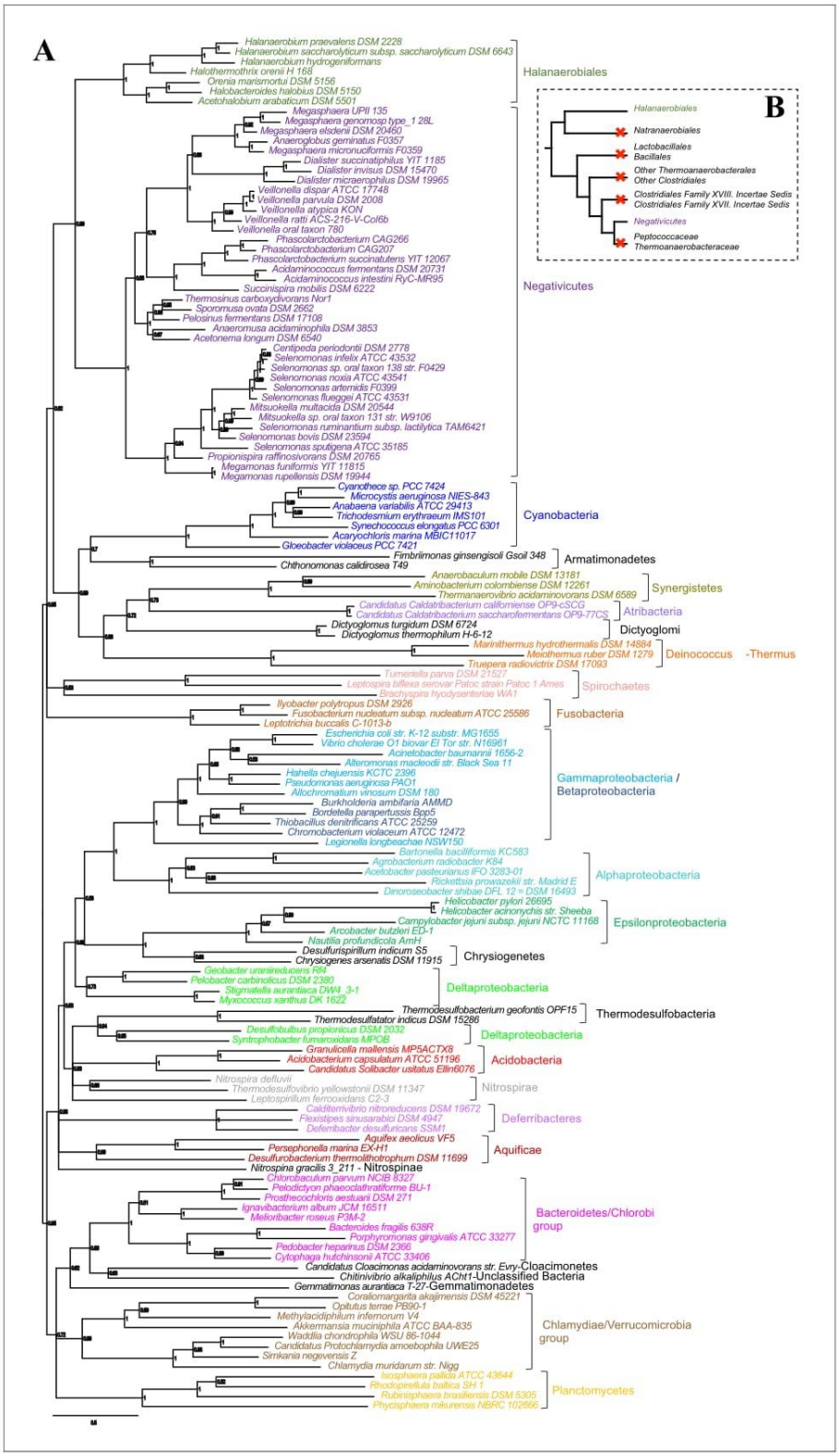

Figure 4. Phylogenetic tree of core LPS components. (**A**) Bayesian phylogeny based on a concatenation of orthologs of the four core components of the LPS biosynthesis pathway (*lpxABCD*), comprising 898 amino acid positions and the CAT+GTR+Γ4 model. Values at nodes represent Bayesian posterior probabilities. The scale bar represents the average number of substitutions per site. For details on analyses, see Materials and methods. (**B**)

Figure 4 continued on next page

Figure 4 continued

Schematic representation of the Firmicutes phylum-level phylogeny from ***Figure 2***, onto which putative losses of the OM are mapped (red crosses). See text for discussion.

DOI: 10.7554/eLife.14589.009

The following figure supplement is available for figure 4:

Figure supplement 1. Results of IC congruence test for the 4 LPS core proteins.

DOI: 10.7554/eLife.14589.010

synthesis and transport. If O-antigen is produced, it is ligated to the lipid A-core by an integral

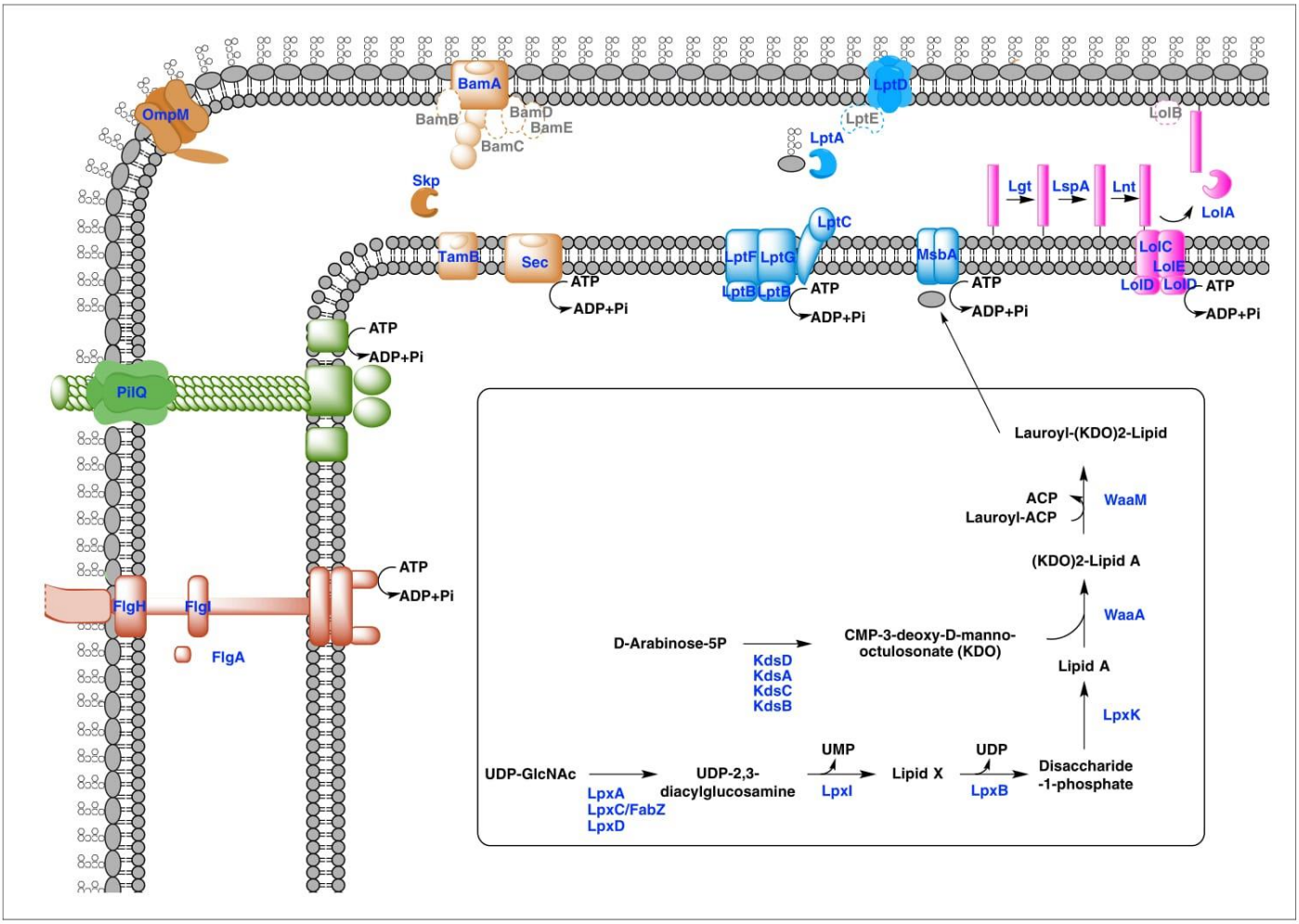

Figure 5. Sketched diagram of inferred characteristics of the diderm Firmicutes cell envelope. The main processes discussed in the text are shown schematically. Components that were not detected in the genomes of Negativicutes and Halanaerobiales are indicated with a dashed outline and grey font.

DOI: 10.7554/eLife.14589.011

The following figure supplements are available for figure 5:

Figure supplement 1. Flagellar gene cluster of Negativicutes and Halanaerobiales.

DOI: 10.7554/eLife.14589.012

Figure supplement 2. Genomic context of the genes coding for flagellar rings in Halanaerobiales and Negativicutes.

DOI: 10.7554/eLife.14589.013

Figure supplement 3. Structure of the main Type IV pilus cluster in Negativicutes and Halanaerobiales.

DOI: 10.7554/eLife.14589.014

membrane ligase, generating 'smooth' LPS. Then, the transport of LPS to the OM is carried out by a dedicated system (the Lpt pathway) and is independent from O-antigen presence (**Wang et al., 2010**).

Both Negativicutes and Halanaerobiales have a complete set of genes for the synthesis of the Lipid A moiety and inner core of LPS (**Figure 5**). These are clustered on their genomes (*lpxACDIBK/waaM/waaA/kdsABCD*) (**Figure 3** and **Supplementary file 1**). The lack of *lpxH* and the presence of *lpxI* suggests that they use a recently-described alternative route for lipid A-core synthesis found in members of various bacterial lineages (**Metzger and Raetz, 2010**). Moreover, the *lpx* gene cluster also includes a *fabZ* homologue, which encodes an enzyme involved in a key step of fatty acid synthesis and is also present in monoderm Firmicutes (**Parsons and Rock, 2013**). Similarly to what is observed in *E. coli*, this *fabZ* homologue shows a conserved position on the chromosome next to *lpxC*, which catalyzes the first committed step in lipid A biosynthesis and is subject of tight regulation by the protease FtsH (**Führer et al., 2006**). Because FabZ acts on a substrate that is shared with the lipid A pathway, its conserved genomic proximity with LpxC is a strong indication that these two proteins interact in diderm Firmicutes, likely in a common process that regulates the phospholipid/LPS balance of the OM as described in *E. coli* (**Klein et al., 2014**; **Ogura et al., 1999**).

Once synthesized, the lipidA-core is translocated through the IM by the flippase MsbA, which is composed of a typical architecture including an N-terminal ABC-transporter transmembrane domain and a C-terminal ATP-binding cassette domain (**Ruiz et al., 2009**). We found homologues with the same domain arrangement within the LPS synthesis gene cluster (**Figure 3** and **Supplementary file 1**). Although these are part of the very large protein family of ABC transporters that is also present in monoderm Firmicutes, their genomic location suggests that they are likely *bona fide* MsbA functional equivalents in diderm Firmicutes (**Figure 5**).

LPS is then matured and transported across the periplasm to the OM via the Lpt pathway (**Greenfield and Whitfield, 2012**; **Polissi and Sperandeo, 2014**; **Ruiz et al., 2009**). LPS is extracted from the IM by LptC and the LptFGB transporter, then mobilized to the OM by the chaperone LptA, and finally assembled into the OM by LptDE. In all Negativicutes and Hanaerobiales, we identified a conserved Lpt four-gene cluster next to the LPS synthesis genes (**Figure 3** and **Supplementary file 1**). This includes a homologue of LptB, a single homologue of LptF/G, plus two proteins of ~200aa containing OstA domains that may represent LptA and LptC (**Figure 5**). Apart from LptB, which belongs to the large P-loop-NTPase superfamily, none of the remaining putative components are present in monoderm Firmicutes. However, we could not find any clear homologues of the OM components LptD/E in the gene cluster or elsewhere in the genomes. Either these are too distantly related to being identified by sequence similarity, or Negativicutes and Halanaerobiales employ a non-homologous system to address LPS to the OM.

Concerning O-antigen, no obvious homologues of its synthetic pathway, as well as its transport through the IM and its ligation with the lipidA-core are present in the conserved gene cluster. Moreover, although we could identify some homologues in the genomes of Negativicutes and Halanaerobiales, these are part of very large protein families and are shared with other pathways, making it difficult to assess functional homology and infer with confidence if diderm Firmicutes are able to make smooth LPS.

An ancient bi-functional Bam/Tam machinery?

The genomic locus also encodes homologues of two important components of the *E. coli* Sec-dependent OM protein assembly pathway (**Selkrig et al., 2014**): one copy of the major component of the Bam system (BamA), which is involved in protein translocation and OM assembly, and three copies of the periplasmic chaperone Skp (**Figure 3** and **Supplementary file 1**). However, we could not find any homologues of the associated lipoproteins of the Bam complex (BamBCDE) in the gene cluster nor elsewhere in the genomes of diderm Firmicutes, though it is known that these are not well conserved outside Proteobacteria (**Webb et al., 2012**).

In Proteobacteria, BamA is part of the large Omp85-family with function in protein translocation and OM assembly (**Heinz and Lithgow, 2014**; **Selkrig et al., 2014**). Members of this family have a conserved C-terminal OM-associated surface antigen domain, and polypeptide transport–associated (POTRA) domains of variable number that serve to interact with other proteins and accessory factors (**Webb et al., 2012**). In Proteobacteria, a close paralogue of BamA was recently discovered, called

TamA, which has one POTRA domain only, and was shown to be anchored to the OM and to form a two-pathway system together with the IM protein TamB to promote T5SS autotransporter assembly and secretion (**Selkrig et al., 2012**). Based on the large taxonomic distribution of TamB homologues, it has been recently proposed that this protein was present very early in bacterial evolution, and functioned with BamA, while TamA would have arisen more recently in Proteobacteria through gene duplication (**Heinz et al., 2015**).

We found that all analyzed Negativicutes and Halanaerobiales genomes possess a single BamA homologue, which is part of a gene cluster with a homologue of TamB, and up to three homologues of the Skp chaperone (**Figure 3** and **Supplementary file 1**). Such conserved synteny may indicate functional linkage. Interestingly, the same gene arrangement (TamB/BamA/Skp) has been observed in Spirochaetes (**Selkrig et al., 2012**), as well as in many other bacterial phyla (**Heinz et al., 2015**). The inference of a potential TamB/BamA/Skp system in the ancestor of the Firmicutes strengthens the hypothesis of an ancestral bi-functional role for BamA in both autotransporter secretion (together with TamB) and OMP assembly (together with Skp) (**Figure 5**).

An ancestral system for peptidoglycan anchoring to the OM

The genomic locus also includes multiple homologues of OmpM (**Figure 3** and **Supplementary file 1**), a protein that appears to replace the function of Braun's lipoprotein in Negativicutes (see Introduction). In *S. ruminantium* OmpM results from the fusion of a C-terminal porin domain and an N-terminal S-layer homologous (SLH) domain with an unusual inward orientation towards the periplasm where it drives the correct anchoring of the OM to the peptidoglycan layer *via* specific interactions with the peptidoglycan-covalently-bound cadaverine (**Kalmokoff et al., 2009**; **Kojima et al., 2010**).

We found that Halanaerobiales also contain in the conserved genomic locus up to three proteins with an SLH domain (**Figure 3** and **Supplementary file 1**). These likely represent distant homologues of Negativicutes OmpM and are priority targets for experimental validation to confirm the presence of an OmpM-dependent system for OM anchoring in this deep-emerging lineage of the Firmicutes. Interestingly, there have been early reports of distantly related proteins with a domain arrangement similar to that of diderm Firmicutes OmpM and a proposed similar function in Cyanobacteria (**Hansel et al., 1998**) and in Thermotogae (**Engel et al., 1992**). OmpM-like proteins might therefore represent a widespread and ancient PG-OM anchoring system (**Figure 5**), possibly ancestral to the Braun's lipoprotein-based system.

Monoderm appendages with diderm features

The monoderm-type flagellum has been mostly studied in *B. subtilis* (**Mukherjee and Kearns, 2014**). It is very similar in overall structure and number of components to the flagellum of diderm bacteria, to the exception of the absence of the rings P and L, the first spanning the periplasm and the second spanning the OM, which are coded by *flgH, flgI*, and *flgA* and are totally absent in monoderm flagella. Another difference is the absence of the rod cap and flagellar-specific peptidoglycan hydrolases, which allow insertion in the PG layer in diderm bacteria, and it is unknown how this process takes place in the thick PG layer of *B. subtilis* (**Mukherjee and Kearns, 2014**).

Most Negativicutes and Halanaerobiales strains analyzed in this study are flagellated (**Vos et al., 2009**) and have in fact a complete set of genes coding for the flagellum machinery, most of which embedded in an operon arrangement that is distinctive of *B. subtilis* and other monoderm Firmicutes, in particular the presence of the two rod assembly genes *flgB* and *flgC* (**Figure 5—figure supplement 1**) (**Mukherjee and Kearns, 2014**). This indicates that the flagella of diderm Firmicutes are closely related to those of their monoderm relatives. However, Negativicutes and Halanaerobiales also possess homologues of the three genes coding for the P and L rings typical of diderm flagella, *flgH, flgI*, and *flgA*, organized in a second conserved six-genes cluster that also contains the rod components *flgF, flgG*, and *flgJ* (**Figure 5—figure supplement 2** and **Supplementary file 1**). These data suggest that Negativicutes and Halanaerobiales might possess flagella with chimeric monoderm/diderm features (**Figure 5**). They may represent ancestral motility structures that adapted to loss of the OM in present-day monoderm Firmicutes lineages, possibly through a single excision of the *flgH, flgI*, and *flgA* gene cluster from the conserved genomic locus. Consistently with this hypothesis, we observed that representatives of the monoderm Firmicutes lineages that are most

closely related to Halanaerobiales (i.e. Natranaerobiales: *N. thermophilus*) and Negativicutes (i.e. Peptococcaceae: *Therminicola potens*) still retain an *flgJ* homologue (***Figure 5—figure supplement 1***) that may represent a remnant of the ancestral diderm system for rod insertion, which would have been lost in *B. subtilis* and other monoderm flagellated Firmicutes.

A similar, although less clear-cut, scenario could be inferred for type IV pili. These have very similar components in diderm and monoderm bacteria, to the exception of PilQ (also called secretin), which in diderm bacteria forms a channel that spans the OM and through which the pilus is assembled (***Melville and Craig, 2013***). In Firmicutes, it is unclear how the pilus passes through the thick peptidoglycan layer, and if a channel-forming protein is present (***Melville and Craig, 2013***). Both Negativicutes and Halanaerobiales have a conserved genomic locus that presumably codes for all essential components of a type IV pilus (***Figure 5—figure supplement 3***), consistently with the visualization of a potential pilus structure by microscopy in ***Figure 1***. The gene cluster is similar to those present in their closely related monoderm relatives, to the exception of the inclusion of a homologue of PilQ, the secretin, which is present in all Negativicutes and Hanalerobiales (***Figure 5—figure supplement 3*** and ***Supplementary file 1***). These data suggest that the type IV pili of Negativicutes and Halanaerobiales may pass through the OM by a secretin-based mechanism similar to that of classical diderm bacteria (***Figure 5***).

First suggestion of a Lol system in the Firmicutes

Lipoproteins are a major class of membrane proteins that play important physiological roles and are widespread in bacteria (***Buddelmeijer, 2015***; ***Okuda and Tokuda, 2011***; ***Sutcliffe et al., 2012***; ***Zückert, 2014***). They are assembled via post-translational modifications after translocation through the inner membrane in both diderm and monoderm bacteria. The lipid modification occurs on a 'lipobox' signal near the N-terminus at the cytoplasmic membrane where the pre-lipoprotein diacylglycerol transferase Lgt adds a diacylglycerol moiety. The lipobox signal is then cleaved off by the lipoprotein signal peptidase LspA, which allows anchoring of the lipoprotein to the outer face of the plasma membrane. A second lipoprotein N-acetyltransferase, Lnt, can add a third amide fatty acid onto the N-terminus of lipoproteins. Tri-acylated lipoproteins are preferentially targeted to the OM in diderm bacteria via the Lol machinery. This is a multiprotein system formed by an IM-spanning ABC-like transporter component (LolCDE) that captures lipoproteins from the IM and delivers them to the carrier LolA utilizing ATP hydrolysis. LolA then translocates the lipoprotein to the structurally related LolB, which inserts it into the OM by an unclear mechanism.

While LolCDE are part of the large protein family of ABC transporters, the presence of LolA homologues is indicative of a functional Lol system (***Sutcliffe et al., 2012***). We found no homologues of LolA nor of LolB in any of the Negativicutes genomes, confirming previous reports (***Campbell et al., 2014***). It is possible that Negativicutes do not address lipoproteins to the OM or that they use a still unknown, non-homologous system. In contrast, we found homologues of LolA in each of the analyzed Halanaerobiales genomes, which correlate to the presence of Lnt homologues (***Supplementary file 1***). By looking at the genomic surroundings of these LolA homologues, we identified a conserved gene cluster in most Halanaerobiales genomes, which likely code for a complete Lol system, notably the two ABC transporter permeases (LolC/E) and one ATPase (LolD) (***Supplementary file 1***). We could not find any homologues of LolB in the genomic locus or elsewhere in the Halanaerobiales genomes, although this protein is not well conserved in bacteria (***Sutcliffe et al., 2012***). These analyses suggest that Halanaerobiales may harbor a functional Lol system (***Figure 5***).

New OM-related proteins?

Finally, three genes encoding proteins with unclear annotation display a conserved synteny within the large conserved genomic cluster (***Figure 3*** and ***Supplementary file 1***) and are present in all Negativicutes and Halanaerobiales genomes (***Supplementary file 1***), while absent from monoderm Firmicutes. The first one belongs to COG0816, which is annotated as 'Predicted endonuclease involved in recombination (possible Holliday junction resolvase in Mycoplasmas and *B. subtilis*)'. The second one belongs to COG4372, which is annotated as 'Uncharacterized conserved protein, contains DUF3084 domain'. The third one belongs to COG750, which is annotated as "Predicted membrane-associated Zn-dependent proteases 1", and is predicted to be in the OM by the PSORT

prediction software (Materials and methods). These proteins might be involved in OM biogenesis and functioning in diderm Firmicutes and are priority targets for experimental characterization.

Discussion

The origin of the cell envelope has represented one of the most fascinating questions in evolutionary biology since decades, and has been widely discussed in the literature (**Blobel, 1980**; **Cavalier-Smith, 1987**, **2006**; **Errington, 2013**; **Griffiths, 2007**; **Gupta, 2011**; **Koch, 2003**; **Vollmer, 2012**). The main issue mostly revolves around the question of how and when an OM originated in Bacteria, and whether monoderm cell envelopes predate diderm cell envelopes or instead derived from them. The complexity of the diderm cell envelope with respect to a perceived more 'rudimentary' monoderm type, together with its higher resistance toward antibiotics, are arguments usually put forward to propose that the OM is a relatively late invention in Bacteria (**Koch, 2003**; **Gupta, 2011**). However, it is now evident that diderm phyla outnumber monoderm ones, and span a large fraction of bacterial diversity, including early emerging lineages (**Sutcliffe, 2010**; **Errington, 2013**). Unfortunately, the evolutionary relationships among monoderm and diderm bacterial phyla are presently unclear, and do not allow to clarify OM origins. In this respect, the existence of a major bacterial phylum –the Firmicutes- including both diderm and monoderm lineages, and whose evolutionary relationships can be resolved, provides a unique opportunity to address the issue.

Our results provide support for the hypothesis that the Firmicutes are ancestrally diderm, and that the monoderm envelope is a derived cell structure that originated through OM loss, at least in this phylum. Although previously suggested (**Tocheva et al., 2011**), the inclusion of both Halanaerobiales and Negativicutes in our analysis strengthens and extends this scenario. Our robust phylogeny of the Firmicutes indicates that Halanaerobiales and Negativicutes form two distinct lineages, each related to different monoderm relatives. This allows inferring that the OM was lost from three to five times independently in the Firmicutes, and is therefore not a unique event in the history of Bacteria that would have led to all present-day monoderm lineages, as proposed earlier (**Cavalier-Smith, 2006**). The deep branching of Halanaerobiales and the still limited genomic coverage for this group makes them a priority target for further exploration as their cell envelopes may retain ancestral characters.

Our results confirm that the LPS-OMs of Negativicutes and Halanaerobiales are homologous structures that share a common origin and are evolutionarily related with those of other classically diderm bacteria, therefore excluding convergence. Indeed, we show that the core enzymatic apparatus for producing LPS is even more widespread than previously thought, and that the LPS-OM is an ancient feature that emerged once and was largely inherited vertically during bacterial evolution, including in Halanaerobiales and Negativicutes. This is unusual for cytosolic enzymes, and underlines the key importance of maintaining cell-envelope function. Clearly, the availability of genomic data from an ever-wider sampling of bacterial diversity is sensibly changing perspective on the evolution of bacterial cell envelopes. For example, by revealing the presence of an LPS-OM in the ancestor of Deinococcus/Thermus we show that this phylum does not represent a monoderm-to-diderm intermediate (**Gupta, 2011**), but rather that LPS was lost in some of their members, a process similar to what likely occurred in Thermotogae (**Cavalier-Smith, 2006**). This indicates that, although having an OM is surely advantageous in certain conditions, the diderm cell envelope is a flexible structure that can be modified/simplified during evolution.

Although the presence of a large cluster coding for key OM functions might be seen as supporting the hypothesis of an acquisition of the OM of diderm Firmicutes via genetic transfer from a diderm bacterium (**Mavromatis et al., 2009**; **Campbell et al., 2014**), our data weaken this hypothesis. Given the distinct branching of the two diderm lineages in the Firmicutes phylogeny, and the pattern of LPS inheritance similar to all other diderm bacterial phyla, the gene transfer hypothesis would imply a complex scenario consisting of two independent transfers of a very large genomic region, a first one to the ancestor of Halanaerobiales or to the ancestor of Negativicutes, and a second one between these two ancestors, which would have had to coexist at the same time and in the same environment. The sudden acquisition of an OM has been already discussed as mechanistically complicated (**Cavalier-Smith, 2006**). This would have in fact required the dramatic modification of an originally monoderm cell envelope through the concerted acquisition of several complex systems at once, not to mention the replacement of the native inner membrane components of these

systems or their coordination with the newly acquired ones. Moreover, not all OM systems involved in OM biogenesis and function in Negativicutes and Halanaerobiales are part of the gene cluster, and their detailed annotation suggests that their cell envelopes share characteristics with deep-emerging bacterial phyla (e.g. the OmpM porin instead of Braun's lipoprotein for OM tethering), and present a peculiar combination of monoderm/diderm features (e.g. flagella, pili). The OM gene cluster may therefore represent an ancestral genomic locus for OM–related functions, and it becomes essential to obtain further experimental characterization of the OM of Halanaerobiales and Negativicutes, as well as many other poorly explored deep emerging bacterial phyla. Nevertheless, we have analyzed here only a few OM systems shared between Halanaerobiales and Negativicutes in order to infer the nature of the ancestral diderm cell envelope in the Firmicutes. There are surely additional components that are lineage-specific and may have been acquired from diderm bacteria thriving in the same environment. This is an important future area of investigation, as it could inform on how the presence of an OM in a Firmicutes background may have helped adaptation to specific niches, including the human environment.

By which process the ancestral OM would have been lost multiple times independently in the majority of present-day Firmicutes remains to be understood. It has been proposed that a spontaneous mutation leading to hypertrophy of the peptidoglycan layer would be sufficient to transform a diderm into a monoderm, through disruption of the attachment of the OM, leading to its loss (***Cavalier-Smith, 2006***). Alternatively, we speculate that mutations may have affected the ancestral OmpM, causing a de-regulation of OM-PG attachment. This transition may have been made easier during the process of sporulation, where an OM is transiently formed and lost when the vegetative cell matures. ***Tocheva et al. (2011)*** proposed indeed that the OM might have been lost in the Firmicutes to increase sporulation and germination efficiency (***Tocheva et al., 2011***). A link between OM loss and sporulation may explain why there is no current evidence of monoderm lineages within other diderm phyla that do not sporulate. Further genomic and experimental data on the closest monoderm relatives of Negativicutes and Halanaerobiales might provide key information on the process of OM loss.

Our results suggest that the cell envelopes of diderm Firmicutes might be the remnants of an ancient type of cellular structure, adding up to the ones found in the major diderm bacterial phyla. Moreover, they seem to have ancestral and simpler cell envelope systems with respect to the well-studied Proteobacteria. Halanaerobiales and Negativicutes are therefore promising new experimental models that will provide precious insights into the processes that have shaped the diversity not only of diderm cell envelopes, but also of monoderm ones.

Materials and methods

Ultrastructural analysis

Negativicutes strain *Megamonas rupellensis* DSM 19944^T was grown anaerobically at 37°C to late exponential phase on TGY broth (w/v; 3% tryptone, 2% yeast extract, 0.5% glucose, 0.05% L-cysteine hydrochloride) as described previously (***Chevrot et al., 2008***). Halanaerobiales strain *Halanaerobium saccharolyticum subsp. saccharolyticum* DSM 6643^T was grown anaerobically to late exponential phase at 37°C on a synthetic medium as described in (***Zhilina et al., 2012***). The ultrastructure of the respective bacterial strains was determined via transmission electron microscopy (TEM) following high-pressure freezing, freeze substitution, plastic embedding and ultrathin sectioning of the samples. Due to a relatively long transportation time to the high-pressure freezer, cells were pre-fixed with 2% glutaraldehyde. Afterwards, they were centrifuged for 10 min at 14.000 x g, the supernatant was discarded and the resulting pellet was resuspended in 50µl growth medium. From this cell suspension, 2µl were high-pressure frozen and freeze substituted as described in (***Peschke et al., 2013***). For substitution, acetone containing 0.2% OsO_4, 0.25% uranyl acetate and 5% (vol/vol) H_2O was used. Embedding of the cells, sectioning and post-staining was carried out as described in (***Rachel et al., 2010***). For negative staining of bacteria, 2% uranyl acetate was used for contrast enhancement following pre-fixation with 2% glutaraldehyde (***Rachel et al., 2010***). Finally, transmission electron microscopy was performed either on a JEOL JEM 2100, operated at 120 kV in combination with a fast-scan 2k x 2k CCD camera F214 (TVIPS, Gauting, Germany) for negatively

stained samples or on a Zeiss EM 912 equipped with an integrated OMEGA energy filter and operated at 80 kV in the zero-loss mode for ultrathin sections.

Phylogenetic analyses

We assembled a local databank of 205 complete genomes from a wide representative sampling of Firmicutes, including 38 Negativicutes and 7 Halanaerobiales genomes available at the beginning of this analysis (***Supplementary file 1***). Exhaustive HMM-based homology searches were carried out on this genome databank by using the HMMER package (***Johnson et al., 2010***) and as queries the HMM profiles of the complete set of 54 bacterial ribosomal proteins from the Pfam 29.0 database (http://pfam.xfam.org, ***Finn et al., 2016***). Additional searches with tblastn (***Altschul et al., 1997***) were used to identify eventually misannotated homologues in some genomes. Because it is unclear which bacterial phylum is closest to the Firmicutes, we included as outgroup 13 taxa representatives of eight major bacterial phyla (2 Actinobacteria; 2 Cyanobacteria; 1 Deinococcus; 2 Proteobacteria; 1 Spirochaetes; 3 Flavobacteria/Bacteroidetes/Chlorobi; 2 Plactomyces/Chlamydia). Seven ribosomal proteins (S2, S4, S14, S21, L25, L30, L33) that were absent from >50% of the considered genomes or had paralogous copies making difficult the identification of orthologues were discarded from the analysis. The remaining 47 single protein data sets were aligned with MUSCLE v3.8.31 (***Edgar, 2004***) with default parameters, and unambiguously aligned positions were selected with BMGE 1.1 (***Criscuolo and Gribaldo, 2010***) and the BLOSUM30 substitution matrix.

Trimmed datasets were concatenated by allowing a maximum of 11 missing sequences per taxon into a large character supermatrix (218 taxa and 5551 amino acid characters). PhyloBayes v3.3b (***Lartillot et al., 2009***) was used to perform Bayesian analysis using the evolutionary model CAT +GTR+Γ_4. Two independent chains were run until convergence, assessed by evaluating the discrepancy of bipartition frequencies between independent runs. The first 25% of trees were discarded as burn-in and the posterior consensus was computed by selecting one tree out of every two . A Maximum likelihood (ML) tree was also calculated from the ribosomal protein concatenate with PhyML v3.0 (***Guindon et al., 2010***) and the evolutionary model LG+Γ_4 (***Le and Gascuel, 2008***) as suggested by the Akaike Information Criterion implemented in ProtTest 3 (***Darriba et al., 2011***). Branch supports were estimated by standard nonparametric bootstrap based on 100 replicates.

In order to assess whether the ribosomal proteins carried a congruent phylogenetic signal, we compared each of the 47 corresponding individual gene trees with the Bayesian ribosomal protein concatenate tree by using the recently proposed 'Internode Certainty' (IC) criterion, which measures the existence of statistically supported conflicting splits between trees (***Kobert et al., 2016***). ML phylogenetic trees of individual genes were inferred by IQ-TREE v1.3.12 (***Nguyen et al., 2015***) with evolutionary model selected by optimizing the Akaike information criterion. In order to minimize the negative impact on IC estimation of the large irresolution within most of the single gene trees caused by the small number of aligned characters, all branches displaying <70% bootstrap support were collapsed. IC values were then estimated by RAxML 8.2.8 (***Stamatakis, 2006***) and reported on the concatenate tree.

Significance of 12 alternative tree topologies was assessed by the approximately unbiased (AU) test (***Shimodaira, 2002***). Each alternative topology was obtained by moving specific nodes on the Bayesian concatenate ribosomal protein tree by using Seaview v4.6 (***Gouy et al., 2010***). For each tree topology, log-likelihood per character was estimated by PhyML v3.0 with the evolutionary model LG+Γ_4. In order to estimate the AU test p-values associated to each topology, the resulting data were processed with CONSEL v0.20 (***Shimodaira and Hasegawa, 2001***) with default parameters.

For the LPS core gene analysis, homologues were searched by using Pfam HMM profiles for LpxA, LpxB, LpxC, and LpxD. The same approach as the one described above for ribosomal proteins was used to assemble a 4-gene supermatrix of 898 unambiguously aligned amino acid characters, which was analysed by PhyloBayes with the evolutionary model CAT+GTR+Γ_4. Congruence among the four markers was assessed by the IC criterion as described above.

Protein annotation

Given the small number of markers analyzed and their frequently limited conservation at the sequence level, we followed a semi-manual procedure for annotation based on a combination of

profile-based homology searches, protein domain identification, and genomic synteny. Profile-based homology searches of specific OM markers were performed by using the HMMER package (**Johnson et al., 2010**). Initial searches were conducted by HMM on the local Firmicute genome databank by using standard Pfam domain models corresponding to a given protein of interest. The top-scoring hits were used to build new HMM models and perform a novel search. This process was iterated until no new hits were found. The resulting homologues were aligned and manually inspected in order to confirm homology and to eliminate divergent, partial or non-homologous sequences. Additional searches with tblastn (**Altschul et al., 1997**) were used to identify eventually misannotated homologues in some genomes. Protein domains were inspected by querying the Conserved Domain Database (CDD) at NCBI (**Marchler-Bauer et al., 2015**). Genomic synteny was investigated using the interactive web-based visualization tool SyntTax (**Oberto, 2013**). Local genomic alignments were generated using EasyFig (**Sullivan et al., 2011**) with a BLAST cutoff E-value of 0.0001. For localization prediction, we used the PSORT v3.0 server (http://www.psort.org/psortb/, **Yu et al., 2010**) with default settings for Gram-negative Bacteria. Protein families were assigned to Clusters of Orthologous Groups by searching the COG database, which was downloaded from the NCBI FTP server (ftp://ftp.ncbi.nih.gov/pub/wolf/COGs/) (**Tatusov et al., 2000**).

Note

While this manuscript was in the last phase of revision, Tocheva et al. published a Perspective paper (Tocheva, EI, Ortega, DR & G Jensen. 2016. Sporulation, bacterial cell envelopes and the origin of life. *Nature Reviews Microbiology* 14, 533-542. doi:10.1038/nrmicro.2016.85). It extends the discussion of their previous hypothesis (**Tocheva et al., 2011**) by focusing on the origin of the outer membrane, and prompts for further genomic and evolutionary analysis, which is timely addressed in the present work.

Acknowledgements

LA was the recipient of a postdoctoral fellowship from the Investissement d'Avenir Grant "Ancestrome" (ANR-10-BINF-01-01). DP is a scholar in the Pasteur-Paris University International PhD program. AK wishes to thank Uwe-G Maier for allocation of the EM facility in Marburg and Marion Debus for technical assistance. We thank the Collection of Bacteria of the Institut Pasteur for kindly providing the two strains used in this work for electron microscopy, and the Centre Informatique pour la Biologie (CIB, pole cluster) for providing computing facility. We also wish to thank Patrick Forterre for discussion during the preparation of this work.

Additional information

Funding

Funder	Grant reference number	Author
Agence Nationale de la Recherche	ANR-10-BINF-01-01	Céline Brochier-Armanet Simonetta Gribaldo

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

LCSA, DP, AK, AC, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; BD, CB-A, CB, Analysis and interpretation of data, Drafting or revising the article; SG, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article

Author ORCIDs

Simonetta Gribaldo, http://orcid.org/0000-0002-7662-021X

Additional files

Supplementary files

• Supplementary file 1. Full distribution and accession numbers of the protein families discussed in the text.

DOI: 10.7554/eLife.14589.015

• Supplementary file 2. Full distribution and accession numbers of the four core LPS genes in a local databank of 121 complete genomes representatives of all major bacterial phyla as discussed in the text. NA indicates that the taxon is unlisted in the NCBI Taxonomy database, while NH indicate that no homologues were found based on HMM search.

DOI: 10.7554/eLife.14589.016

Major datasets

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
Luisa CS Antunes, Daniel Poppleton, Andreas Klingl, Alexis Criscuolo, Bruno Dupuy, Céline Brochier-Armanet, Christophe Beloin, Simonetta Gribaldo	2016	Data from: Phylogenomic analysis supports the ancestral presence of LPS-outer membranes in the Firmicutes	http://dx.doi.org/10.5061/dryad.48m7c	Available at Dryad Digital Repository under a CC0 Public Domain Dedication

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**:3389–3402. doi: 10.1093/nar/25.17.3389

Antunes LCS, Poppleton D, Klingl A, Criscuolo A, Dupuy B, Brochier-Armanet C, Beloin C, Gribaldo S. 2016. Data from: Phylogenomic analysis supports the ancestral presence of LPS-outer membranes in the Firmicutes. *Dryad Digital Repository* . doi: 10.5061/dryad.48m7c

Arrieta MC, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, Kuzeljevic B, Gold MJ, Britton HM, Lefebvre DL, Subbarao P, Mandhane P, Becker A, McNagny KM, Sears MR, Kollmann T, Mohn WW, Turvey SE, Finlay BB, CHILD Study Investigators. 2015. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Science Translational Medicine* **7**:307ra152. doi: 10.1126/scitranslmed.aab2271

Bladen HA, Mergenhagen SE. 1964. Ultrastructure of Veillonella and morphological correlation of an outer membrane with particles associated with endotoxic activity. *Journal of Bacteriology* **88**:1482–1492.

Blobel G. 1980. Intracellular protein topogenesis. *PNAS* **77**:1496–1500. doi: 10.1073/pnas.77.3.1496

Buddelmeijer N. 2015. The molecular mechanism of bacterial lipoprotein modification–how, when and why? *FEMS Microbiology Reviews* **39**:246–261. doi: 10.1093/femsre/fuu006

Campbell C, Sutcliffe IC, Gupta RS. 2014. Comparative proteome analysis of *Acidaminococcus intestini* supports a relationship between outer membrane biogenesis in Negativicutes and Proteobacteria. *Archives of Microbiology* **196**:307–310. doi: 10.1007/s00203-014-0964-4

Cavalier-Smith T. 1987. The origin of cells: a symbiosis between genes, catalysts, and membranes. *Cold Spring Harbor Symposia on Quantitative Biology* **52**:805–824. doi: 10.1101/SQB.1987.052.01.089

Cavalier-Smith T. 2006. Rooting the tree of life by transition analyses. *Biology Direct* **1**:19. doi: 10.1186/1745-6150-1-19

Cayol JL, Ollivier B, Patel BK, Prensier G, Guezennec J, Garcia JL. 1994. Isolation and characterization of *Halothermothrix orenii* gen. nov., sp. nov., a halophilic, thermophilic, fermentative, strictly anaerobic bacterium. *International Journal of Systematic Bacteriology* **44**:534–540. doi: 10.1099/00207713-44-3-534

Chevrot R, Carlotti A, Sopena V, Marchand P, Rosenfeld E. 2008. *Megamonas rupellensis* sp. nov., an anaerobe isolated from the caecum of a duck. *International Journal of Systematic and Evolutionary Microbiology* **58**:2921–2924. doi: 10.1099/ijs.0.2008/001297-0

Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology* **10**:20–10. doi: 10.1186/1471-2148-10-210

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**:1164–1165. doi: 10.1093/bioinformatics/btr088

Davis JJ, Xia F, Overbeek RA, Olsen GJ. 2013. Genomes of the class Erysipelotrichia clarify the firmicute origin of the class Mollicutes. *International Journal of Systematic and Evolutionary Microbiology* **63**:2727–2741. doi: 10.1099/ijs.0.048983-0

Dawes IW. 1980. Sporulation in evolution. In: Collins J. F, Moseley B. E. B (Eds). *Cellular and Molecular Aspects of Microbial Evolution*. Cambridge: Cambridge University Press. p 85–130

Delwiche EA, Pestka JJ, Tortorello ML. 1985. The Veillonellae: gram-negative cocci with a unique physiology. *Annual Review of Microbiology* **39**:175–193. doi: 10.1146/annurev.mi.39.100185.001135

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**:1792–1797. doi: 10.1093/nar/gkh340

Engel AM, Cejka Z, Lupas A, Lottspeich F, Baumeister W. 1992. Isolation and cloning of Omp alpha, a coiled-coil protein spanning the periplasmic space of the ancestral eubacterium *Thermotoga maritima*. *The EMBO Journal* **11**:4369–4378.

Errington J. 2013. L-form bacteria, cell walls and the origins of life. *Open Biology* **3**:120143. doi: 10.1098/rsob.120143

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* **44**:D279–D285. doi: 10.1093/nar/gkv1344

Führer F, Langklotz S, Narberhaus F. 2006. The C-terminal end of LpxC is required for degradation by the FtsH protease. *Molecular Microbiology* **59**:1025–1036. doi: 10.1111/j.1365-2958.2005.04994.x

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* **27**:221–224. doi: 10.1093/molbev/msp259

Greenfield LK, Whitfield C. 2012. Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydrate Research* **356**:12–24. doi: 10.1016/j.carres.2012.02.027

Griffen AL, Beall CJ, Campbell JH, Firestone ND, Kumar PS, Yang ZK, Podar M, Leys EJ, Podar M, Podar & E M. 2012. Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *ISME Journal* **6**:1176–1185. doi: 10.1038/ismej.2011.191

Griffiths G. 2007. Cell evolution and the problem of membrane topology. *Nature Reviews Molecular Cell Biology* **8**:1018–1024. doi: 10.1038/nrm2287

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**:307–321. doi: 10.1093/sysbio/syq010

Gupta RS. 2011. Origin of diderm (Gram-negative) bacteria: antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes. *Antonie Van Leeuwenhoek* **100**:171–182. doi: 10.1007/s10482-011-9616-8

Hansel A, Pattus F, Jürgens UJ, Tadros MH. 1998. Cloning and characterization of the genes coding for two porins in the unicellular cyanobacterium *Synechococcus* PCC 6301. *Biochimica et Biophysica Acta* **1399**:31–39. doi: 10.1016/S0167-4781(98)00090-6

Heinz E, Lithgow T. 2014. A comprehensive analysis of the Omp85/TpsB protein superfamily structural diversity, taxonomic occurrence, and evolution. *Frontiers in Microbiology* **5**:370. doi: 10.3389/fmicb.2014.00370

Heinz E, Selkrig J, Belousoff MJ, Lithgow T. 2015. Evolution of the translocation and assembly module (TAM). *Genome Biology and Evolution* **7**:1628–1643. doi: 10.1093/gbe/evv097

Hirai J, Yamagishi Y, Kinjo T, Hagihara M, Sakanashi D, Suematsu H, Fujita J, Mikamo H. 2016. Osteomyelitis caused by Veillonella species: Case report and review of the literature. *Journal of Infection and Chemotherapy* **22**:417–420. doi: 10.1016/j.jiac.2015.12.015

Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**:43. doi: 10.1186/1471-2105-11-431

Kalmokoff ML, Austin JW, Cyr TD, Hefford MA, Teather RM, Selinger LB. 2009. Physical and genetic characterization of an outer-membrane protein (OmpM1) containing an N-terminal S-layer-like homology domain from the phylogenetically Gram-positive gut anaerobe *Mitsuokella multacida*. *Anaerobe* **15**:74–81. doi: 10.1016/j.anaerobe.2009.01.001

Klein G, Kobylak N, Lindner B, Stupak A, Raina S. 2014. Assembly of lipopolysaccharide in *Escherichia coli* requires the essential LapB heat shock protein. *Journal of Biological Chemistry* **289**:14829–14853. doi: 10.1074/jbc.M113.539494

Kobert K, Salichos L, Rokas A, Stamatakis A, Rokas A, Rokas & A. 2016. Computing the Internode Certainty and Related Measures from Partial Gene Trees. *Molecular Biology and Evolution* **33**:1606–1617. doi: 10.1093/molbev/msw040

Koch AL. 2003. Were Gram-positive rods the first bacteria? *Trends in Microbiology* **11**:166–170. doi: 10.1016/S0966-842X(03)00063-5

Kojima S, Ko KC, Takatsuka Y, Abe N, Kaneko J, Itoh Y, Kamio Y. 2010. Cadaverine covalently linked to peptidoglycan is required for interaction between the peptidoglycan and the periplasm-exposed S-layer-homologous domain of major outer membrane protein Mep45 in *Selenomonas ruminantium*. *Journal of Bacteriology* **192**:5953–5961. doi: 10.1128/JB.00417-10

Kunisawa T. 2015. Evolutionary relationships of completely sequenced Clostridia species and close relatives. *International Journal of Systematic and Evolutionary Microbiology*. doi: 10.1099/ijsem.0.000638

Lake JA. 2009. Evidence for an early prokaryotic endosymbiosis. *Nature* **460**:967–971. doi: 10.1038/nature08183

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**:2286–2288. doi: 10.1093/bioinformatics/btp368

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution* **21**:1095–1109. doi: 10.1093/molbev/msh112

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Molecular Biology and Evolution* **25**:1307–1320. doi: 10.1093/molbev/msn067

Marchandin H, Teyssier C, Campos J, Jean-Pierre H, Roger F, Gay B, Carlier JP, Jumas-Bilak E. 2010. *Negativicoccus succinicivorans* gen. nov., sp. nov., isolated from human clinical samples, emended description of the family Veillonellaceae and description of Negativicutes classis nov., Selenomonadales ord. nov. and Acidaminococcaceae fam. nov. in the bacterial phylum Firmicutes. *International Journal of Systematic and Evolutionary Microbiology* **60**:1271–1279. doi: 10.1099/ijs.0.013102-0

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Research* **43**:D222–D226. doi: 10.1093/nar/gku1221

Mavromatis K, Ivanova N, Anderson I, Lykidis A, Hooper SD, Sun H, Kunin V, Lapidus A, Hugenholtz P, Patel B, Kyrpides NC. 2009. Genome analysis of the anaerobic thermohalophilic bacterium *Halothermothrix orenii*. *PLoS One* **4**:e4192. doi: 10.1371/journal.pone.0004192

McDonald K, Morphew M, Verkade P, Müller-Reichert T. 2007a. Recent advances in high-pressure freezing: equipment- and specimen-loading methods. *Methods in Molecular Biology* **369**:143–173 . doi: 10.1007/978-1-59745-294-6_8

McDonald K. 2007. Cryopreparation methods for electron microscopy of selected model systems. *Methods in Cell Biology* **79**:23–56. doi: 10.1016/S0091-679X(06)79002-1

Melville S, Craig L. 2013. Type IV pili in Gram-positive bacteria. *Microbiology and Molecular Biology Reviews* **77**:323–341. doi: 10.1128/MMBR.00063-12

Mesbah NM, Hedrick DB, Peacock AD, Rohde M, Wiegel J. 2007. *Natranaerobius thermophilus* gen. nov., sp. nov., a halophilic, alkalithermophilic bacterium from soda lakes of the Wadi An Natrun, Egypt, and proposal of Natranaerobiaceae fam. nov. and Natranaerobiales ord. nov. *International Journal of Systematic and Evolutionary Microbiology* **57**:2507–2512. doi: 10.1099/ijs.0.65068-0

Metzger LE, Raetz CR. 2010. An alternative route for UDP-diacylglucosamine hydrolysis in bacterial lipid A biosynthesis. *Biochemistry* **49**:6715–6726. doi: 10.1021/bi1008744

Mukherjee S, Kearns DB. 2014. The structure and regulation of flagella in *Bacillus subtilis*. *Annual Review of Genetics* **48**:319–340. doi: 10.1146/annurev-genet-120213-092406

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**:268–274. doi: 10.1093/molbev/msu300

Oberto J. 2013. SyntTax: a web server linking synteny to prokaryotic taxonomy. *BMC Bioinformatics* **14**:4. doi: 10.1186/1471-2105-14-4

Ogura T, Inoue K, Tatsuta T, Suzaki T, Karata K, Young K, Su LH, Fierke CA, Jackman JE, Raetz CR, Coleman J, Tomoyasu T, Matsuzawa H. 1999. Balanced biosynthesis of major membrane components through regulated degradation of the committed enzyme of lipid A biosynthesis by the AAA protease FtsH (HflB) in *Escherichia coli*. *Molecular Microbiology* **31**:833–844. doi: 10.1046/j.1365-2958.1999.01221.x

Okuda S, Tokuda H. 2011. Lipoprotein sorting in bacteria. *Annual Review of Microbiology* **65**:239–259. doi: 10.1146/annurev-micro-090110-102859

Opiyo SO, Pardy RL, Moriyama H, Moriyama EN. 2010. Evolution of the Kdo2-lipid A biosynthesis in bacteria. *BMC Evolutionary Biology* **10**:362. doi: 10.1186/1471-2148-10-362

Oren A. 2006. The order Halanaerobiales. In: Schleifer and Stackebrand. *The Prokaryotes: A Handbook on the Biology of Bacteria*. New York: Springer. p 809–822.

Parsons JB, Rock CO. 2013. Bacterial lipids: metabolism and membrane homeostasis. *Progress in Lipid Research* **52**:249–276. doi: 10.1016/j.plipres.2013.02.002

Peschke M, Moog D, Klingl A, Maier UG, Hempel F. 2013. Evidence for glycoprotein transport into complex plastids. *PNAS* **110**:10860–10865. doi: 10.1073/pnas.1301945110

Polissi A, Sperandeo P. 2014. The lipopolysaccharide export pathway in *Escherichia coli*: structure, organization and regulated assembly of the Lpt machinery. *Marine Drugs* **12**:1023–1042. doi: 10.3390/md12021023

Rachel R, Meyer C, Klingl A, Gürster S, Heimerl T, Wasserburger N, Burghardt T, Küper U, Bellack A, Schopf S, Wirth R, Huber H, Wanner G. 2010. Analysis of the ultrastructure of archaea by electron microscopy. *Methods in Cell Biology* **96**:47–69. doi: 10.1016/S0091-679X(10)96003-2

Roush D, Elias D, Mormile M, Dwayne &M.R AE. 2014. Metabolic Capabilities of the Members of the Order Halanaerobiales and Their Potential Biotechnological Applications. *Current Biotechnology* **3**:3–9. doi: 10.2174/2211550103011403031027 41

Ruiz N, Kahne D, Silhavy TJ. 2009. Transport of lipopolysaccharide across the cell envelope: the long road of discovery. *Nature Reviews Microbiology* **7**:677–683. doi: 10.1038/nrmicro2184

Selkrig J, Leyton DL, Webb CT, Lithgow T. 2014. Assembly of β-barrel proteins into bacterial outer membranes. *Biochimica Et Biophysica Acta* **1843**:1542–1550. doi: 10.1016/j.bbamcr.2013.10.009

Selkrig J, Mosbahi K, Webb CT, Belousoff MJ, Perry AJ, Wells TJ, Morris F, Leyton DL, Totsika M, Phan MD, Celik N, Kelly M, Oates C, Hartland EL, Robins-Browne RM, Ramarathinam SH, Purcell AW, Schembri MA,

Strugnell RA, Henderson IR, et al. 2012. Discovery of an archetypal protein transport system in bacterial outer membranes. *Nature Structural & Molecular Biology* **19**:506–510. doi: 10.1038/nsmb.2261

Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**:1246–1247. doi: 10.1093/bioinformatics/17.12.1246

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* **51**:492–508. doi: 10.1080/10635150290069913

Silhavy TJ, Kahne D, Walker S. 2010. The bacterial cell envelope. *Cold Spring Harbor Perspectives in Biology* **2**: a000414. doi: 10.1101/cshperspect.a000414

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**:2688–2690. doi: 10.1093/bioinformatics/btl446

Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**:1009–1010. doi: 10.1093/bioinformatics/btr039

Sutcliffe IC, Harrington DJ, Hutchings MI. 2012. A phylum level analysis reveals lipoprotein biosynthesis to be a fundamental property of bacteria. *Protein & Cell* **3**:163–170. doi: 10.1007/s13238-012-2023-8

Sutcliffe IC. 2010. A phylum level perspective on bacterial cell envelope architecture. *Trends in Microbiology* **18**: 464–470. doi: 10.1016/j.tim.2010.06.005

Tanner AC, Mathney JM, Kent RL, Chalmers NI, Hughes CV, Loo CY, Pradhan N, Kanasi E, Hwang J, Dahlan MA, Papadopolou E, Dewhirst FE. 2011. Cultivable anaerobic microbiota of severe early childhood caries. *Journal of Clinical Microbiology* **49**:1464–1474. doi: 10.1128/JCM.02427-10

Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* **28**:33–36. doi: 10.1093/nar/28.1.33

Tocheva EI, Matson EG, Morris DM, Moussavi F, Leadbetter JR, Jensen GJ. 2011. Peptidoglycan remodeling and conversion of an inner membrane into an outer membrane during sporulation. *Cell* **146**:799–812. doi: 10.1016/j.cell.2011.07.029

van den Bogert B, Erkus O, Boekhorst J, de Goffau M, Smid EJ, Zoetendal EG, Kleerebezem M. 2013. Diversity of human small intestinal *Streptococcus* and *Veillonella* populations. *FEMS Microbiology Ecology* **85**:376–388. doi: 10.1111/1574-6941.12127

Veillon NM, Zuber A. 1898. Recherches sur quelques microbes strictement anaérobies et leur rôle en pathologie. *Archives De Médecine Expérimentale et d'Anatomie Pathologique* **10**:517–545.

Vesth T, Ozen A, Andersen SC, Kaas RS, Lukjancenko O, Bohlin J, Nookaew I, Wassenaar TM, Ussery DW. 2013. Veillonella, Firmicutes: Microbes disguised as Gram negatives. *Standards in Genomic Sciences* **9**:431–448. doi: 10.4056/sigs.2981345

Vollmer W. 2012. Bacterial outer membrane evolution via sporulation? *Nature Chemical Biology* **8**:14–18. doi: 10.1038/nchembio.748

Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer K-H, Whitman W. 2009. *Bergey's Manual of Systemic Bacteriology. Volume 3: The Firmicutes*. New York: Springer-Verlag.

Wang CH, Kan LP, Sun JR, Yu CM, Yin T, Huang TW, Tsai WC, Yang YS. 2015. Empyema caused by *Anaeroglobus geminates*, a case report with literature review. *Infection* **43**:117–120. doi: 10.1007/s15010-014-0679-0

Wang L, Wang Q, Reeves PR. 2010. The variation of O antigens in gram-negative bacteria. *Sub-Cellular Biochemistry* **53**:123–152. doi: 10.1007/978-90-481-9078-2_6

Webb CT, Heinz E, Lithgow T. 2012. Evolution of the β-barrel assembly machinery. *Trends in Microbiology* **20**: 612–620. doi: 10.1016/j.tim.2012.08.006

Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**:1608–1615. doi: 10.1093/bioinformatics/btq249

Yutin N, Galperin MY. 2013. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environmental Microbiology* **15**:2631–2641. doi: 10.1111/1462-2920.12173

Zeikus JG, Hegge PW, Thompson TE, Phelps TJ, Langworthy TA. 1983. Isolation and description of *Haloanaerobium praevalens* gen. nov. and sp. nov., an obligately anaerobic halophile common to Great Salt Lake sediments. *Current Microbiology* **9**:225–233. doi: 10.1007/BF01567586

Zhilina TN, Zavarzin GA, Bulygina ES, Kevbrin VV, Osipov GA, Chumakov KM. 1992. Ecology, physiology and taxonomy studies on a new taxon of *Haloanaerobiaceae, Haloincola saccharolytica* gen. nov., sp. nov. *Systematic and Applied Microbiology* **15**:275–284. doi: 10.1016/S0723-2020(11)80101-3

Zhilina TN, Zavarzina DG, Panteleeva AN, Osipov GA, Kostrikina NA, Tourova TP, Zavarzin GA. 2012. *Fuchsiella alkaliacetigenal* gen. nov., sp. nov., an alkaliphilic, lithoautotrophic homoacetogen from a soda lake. *International Journal of Systematic and Evolutionary Microbiology* **62**:1666–1673. doi: 10.1099/ijs.0.034363-0

Zhou P, Liu J, Merritt J, Qi F. 2015. A YadA-like autotransporter, Hag1 in *Veillonella atypica* is a multivalent hemagglutinin involved in adherence to oral streptococci, *Porphyromonas gingivalis*, and human oral buccal cells. *Molecular Oral Microbiology* **30**:269–279. doi: 10.1111/omi.12091

Zückert WR. 2014. Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. *Biochimica et Biophysica Acta* **1843**:1509–1516. doi: 10.1016/j.bbamcr.2014.04.022

1.2 A New Diderm Lineage Strengthens the OM First Hypothesis for Firmicutes

Although the evidence for diderm first, presented in the previous article, was extremely strong, I was never satisfied with it. The combination of having most of the OM genes residing within a single cluster combined with having only two diderm groups did not completely exclude transfer. This unlikely transfer would have occurred from the first Halanaerobiale to the first Negativicute or vice versa. To exclude this possibility I endeavoured to obtain more diderm Firmicute genomes representing new clades. Within Bergey's Manual (Whitman, 2011) there are many Firmicutes with a tenative OM apart from the Negativicutes and the Halanaerobiales. In an effort to identify more diderm lineages, we sequenced four new genomes in hopes of discovering new clades; unfortunately, none of the genomes had any diderm marker genes. These genomes may represent new types of OM or had imaging artefacts, neither of which aid me in firmly demonstrating that the ancestor of Firmicutes was a diderm and as such will not be discussed here.

Fortunately, a third diderm group was reported in the literature, the Limnochordales. The Limnochordales are a new clade of the Firmicutes with only a single-member *Limnochorda pilosa*. *L. pilosa* was isolated from the sediment of a brackish meromictic lake and found to grow at an optimum temperature 45ºC (Watanabe *et al.*, 2015). It was found to fall near the Symbiobacteriaceae and family XVII Incertae Sedis in the class Clostridia by 16s phylogeny and this was further confirmed when the genome was sequenced. When the genome was sequenced and the electron microscopy was performed, the authors discovered that the bacteria could sporulate and appeared to be diderm (Fukui *et al.*, 2016). Additionally, they found the presence of many diderm marker genes, however, they did not go into depth. These bacteria may represent a third diderm clade of Firmicutes, but further analysis was necessary.

I started by repeating all of the steps from the previous analysis however, I could not get a proper phylogenetic position for *L. pilosa* as it moved around the tree depending on the method. As such, I needed more Limnochordale genomes to improve the resolution of the tree. I assembled three new genomes from metagenomic samples and this allowed proper construction of phylogenies, both LPS and ribosomal.

I present to you a draft of the article we hope to submit as an eLIFE advance in research in the coming weeks.

Limnochorda: a Third Diderm Lineage Strengthens Diderm Firmicute Ancestry

Daniel I Poppleton[1], Najwa Taib[1], Guillaume Borrel[1], Christophe Beloin[2], Simonetta Gribaldo[1]*

[1]Unité de Biologie Moléculaire du Gène chez les Extrêmophiles, Département de Microbiologie, Institut Pasteur, Paris, France
[2]Unité de Génétique des Biofilms, Département de Microbiologie, Institut Pasteur, Paris, France

Abstract Among the major transitions in the evolutionary history of Bacteria is that between cell envelopes with one membrane (monoderm or Gram-positive) or two membranes (diderm or Gram-negative). Recently, we studied the phylogenetic placement and genome content of the Negativicutes and the Halanaerobiales, two clades of the Firmicutes that possess outer membranes. We proposed that the ancestor of this major bacterial phylum was a diderm and that outer membranes were lost multiple times independently to give rise to the monoderm cell envelope architecture (Antunes et al., 2016). The alternative hypothesis of an ancient horizontal gene transfer between the Negativicutes and the Halanaerobiales could not be totally excluded. Here, we have updated our analyses to include a third diderm Firmicute lineage, the Limnochordales, which was recently reported from anaerobic lake sediments. We show that Limnochordales represent a distinct branch from Halanaerobiales and Negativicutes. The genomic analysis highlights common features with the other two diderm Firmicutes lineages, but also interesting peculiarities that shed light on the evolution of outer membranes in Bacteria. The existence of a third diderm lineage in the Firmicutes weakens the hypothesis of an ancient horizontal gene transfer and infers even more independent OM losses in this phylum than previously thought. We anticipate that additional diderm lineages may be found in this major bacterial phylum.

Introduction

The cell envelope is one of the most ancient features of life; yet, most aspects of its evolutionary history remain obscure. The phylogenetic relationships among monoderm and diderm bacterial phyla are ill-resolved and the details of such transition have been elusive. In this respect, the existence of both monoderm and diderm lineages within the same phylum represents a unique opportunity to clarify this issue. This is the case of the Firmicutes, which include two clades that display outer membranes, the Negativicutes and the Halanaerobiales (Whitman, 2011).

We recently demonstrated that the Negativicutes and Halanaerobiales form two phylogenetically distinct lineages, each close to different monoderm relatives (Antunes et al., 2016). In addition, phylogenetic analysis of the core biosynthetic LPS genes showed that these were inherited vertically in diderm Firmicutes, as in most diderm bacterial phyla. Annotation of OM markers in the genomes of Halanaerobiales and Negativicutes showed that many of these genes are clustered. Furthermore, it indicated that these two lineages display a puzzling combination of monoderm and diderm features. We have recently confirmed these bioinformatics predictions by producing an OM proteome for the model diderm Firmicute *Veillonella parvula* (Poppleton et al., 2017). We put forward the hypothesis that the diderm envelope of Negativicutes and Halanaerobiales is an ancestral characteristic of the Firmicutes that was retained only in these two lineages, while it was lost multiple times

independently during the diversification of this phylum to give rise to the classical Gram-positive cell envelope architecture. An alternative, although less likely, a scenario wherein a first horizontal gene transfer of the genetic determinants for an OM from a diderm bacterium to the ancestor of Halanaerobiales or the ancestor of Negativicutes, followed by a second transfer between these two ancestors.

Recently, the first member of a new lineage belonging to the Firmicutes, *Limnochorda pilosa*, was isolated from a brackish meromictic lake (Watanabe et al., 2015). On the basis of 16S phylogeny, it was found to be distantly related to the family Symbiobacteriaceae and family XVII Incertae Sedis in the class Clostridia (Watanabe et al., 2015). The genome of *L. pilosa* was released the year after, revealing the presence of classical OM markers, consistent with a ultrastructural evidence for a diderm cell envelope, and a partial set of sporulating genes, a defining characteristic of the Firmicutes (Fukui et al., 2016). The existence of an additional diderm lineage in the Firmicutes provides a great opportunity to test our hypothesis. Here, we have obtained additional genomic data from members of Limnochordales by probing available metagenomics databases and included them in our analyses. Phylogenetic analysis and annotation of OM markers further supports the scenario of a diderm ancestor of the Firmicutes and provides important insights into the biogenesis, functioning, and evolution of the OM in this phylum, as well as all Bacteria.

Results

Assembly of three new genomes from uncultured Limnochordales representatives

Table 1. Characteristics of the new *Limnochorda* genomes obtained from metagenomes. A local databank of all metagenomes available on JGI IMG (Chen et al., 2017) was queried using the RNA polymerase β' of *L. pilosa* (BAS29096.1) by BLAST (Camacho et al., 2009). Phylogenetic analysis of the top hits confirmed that they belonged to strains related to *L. pilosa*. The corresponding metagenomes were therefore gathered to reconstruct the genomes of these strains by using tetramer frequency was calculated (Dick et al., 2009) and binning with emergent self-organizing map (ESOM) (Ultsch and Mörchen, 2005). The resulting genomes were checked for completeness, heterogeneity, and contamination with CheckM (Parks et al., 2015).

Names	Limnochorda Metagenomes	Gold Project ID	Completion %	Contamination %	heterogeneity %	Bp	Contigs	Genes	GC %
Limnochorda sp. B1P2	Biogas fermentation Plant 1 DNA2	Gp0056816	97.46	9.36	23.81	523275	54	3024	47.5
Limnochorda sp. B4P1	Biogas fermentation Plant 4 DNA1	Gp0056815	86.76	6.9	54.55	1848866	246	2007	57.9
None Given	Biogas fermentation Plant 1 DNA1	Ga0079224	12.83	0	0		-	-	-
Limnochorda sp. UASM	Agricultural soil microbial communities from Utah to study Nitrogen management	Gp0056840	54.6	0	0	1406273	173	1427	63.1
Limnochorda pilosa HC45			100	0	0	3817704	Complete Genome	3432	69.7

In order to enrich genome data from the Limnochordales, we queried available metagenomics databases by using as seed the RNA polymerase β' sequence of *L. pilosa*. Four metagenomes were identified that contained related sequence homologs clustering with *L. pilosa* in preliminary phylogenetic trees and therefore belonging to candidate Limnochordale strains. Three of these were from an ethanol producing bioreactor, and the fourth was obtained from agricultural soil samples (Table 1). By binning and clustering these metagenomes, four Limnochordales genomes were reconstructed and checked for contamination, completeness, and heterogeneity. Two

of them (*L.* sp. B1P2 & *L.* sp. B4P1) were nearly complete and had low contamination, one (*L.* sp. UASM) was 54 % complete, while the fourth (Ga0079224) proved too incomplete and was excluded from further study.

Limnochordales have a distinct origin from Negativicutes and Halanaerobiales

To analyze the placement of the Limnochordales within the Firmicutes and their evolutionary relationships with the other two diderm lineages, we built a phylum-level phylogenetic tree based on a concatenation of ribosomal proteins and 229 representative taxa, including some new lineages whose genome became available recently (M&M). Only three Limnochordale genomes could be utilized, as *L. sp.* UASM had only less than 50 % of the ribosomal proteins. The Bayesian tree was well resolved at most nodes and largely consistent with our previous analysis, to the exception of two deep and poorly supported nodes concerning the yet unresolved monophyly or paraphyly of Clostridia with respect to Bacilli (Figure 1). The Limnochordales form a distinct branch from the Negativicutes and Halanaerobiales and fall within the Clostridia forming a robustly supported cluster with the monoderm families Symbiobacteriaceae and XVII Incertae Sedis, a position consistent with previous 16S rDNA analysis (Fukui *et al.*, 2016).

From this analysis, we also managed to assign a robust phylogenetic placement for the Tissierella, a new clade of Firmicutes recently sequenced from human samples and including emerging opportunistic pathogens (Alauzet *et al.*, 2014). Also of note is the position of *Dethiobacter alkaliphilus* as the first sequenced member of a new deep branching lineage in the Natranaerobiales/Halanaerobiales clade (Figure 1). This placement is consistent with the fact that these lineages share similar environmental distribution in alkaline, thermophilic, and halophilic environments. Further exploration of these deep clades in the Firmicutes will be key to understand the very emergence of this important bacterial phylum.

Annotation of OM markers in Limnochordales identifies shared and unique traits.

The existence of a third diderm lineage provides an exceptional opportunity to gain further insights into the biogenesis and function of the OM in the Firmicutes. Therefore, we proceeded to identify and annotate putative OM markers in the genome of L. pilosa and the three most complete reconstructed genomes (Figure 2). L. pilosa possess a large genetic cluster that contains many genes responsible for OM biosynthesis and function (Figure 2), similar to the one that we previously identified in the Negativicutes and Halanaerobiales (Antunes et al., 2016). Within this cluster are the genes responsible for the OM rings of diderm-type flagella, LPS biosynthesis and transport, and β-barrel synthesis machinery. Additionally, it has one or multiple copies of the specific type of OM attachment typical of Negativicutes and Halanaerobiales (OmpM) that is distinct from Braun's lipoprotein (Antunes et al., 2016). Although the Limnochordales assembled from metagenomes did not have a whole contig containing the entire genetic locus, this cluster may be conserved in the other members of this lineage as well.

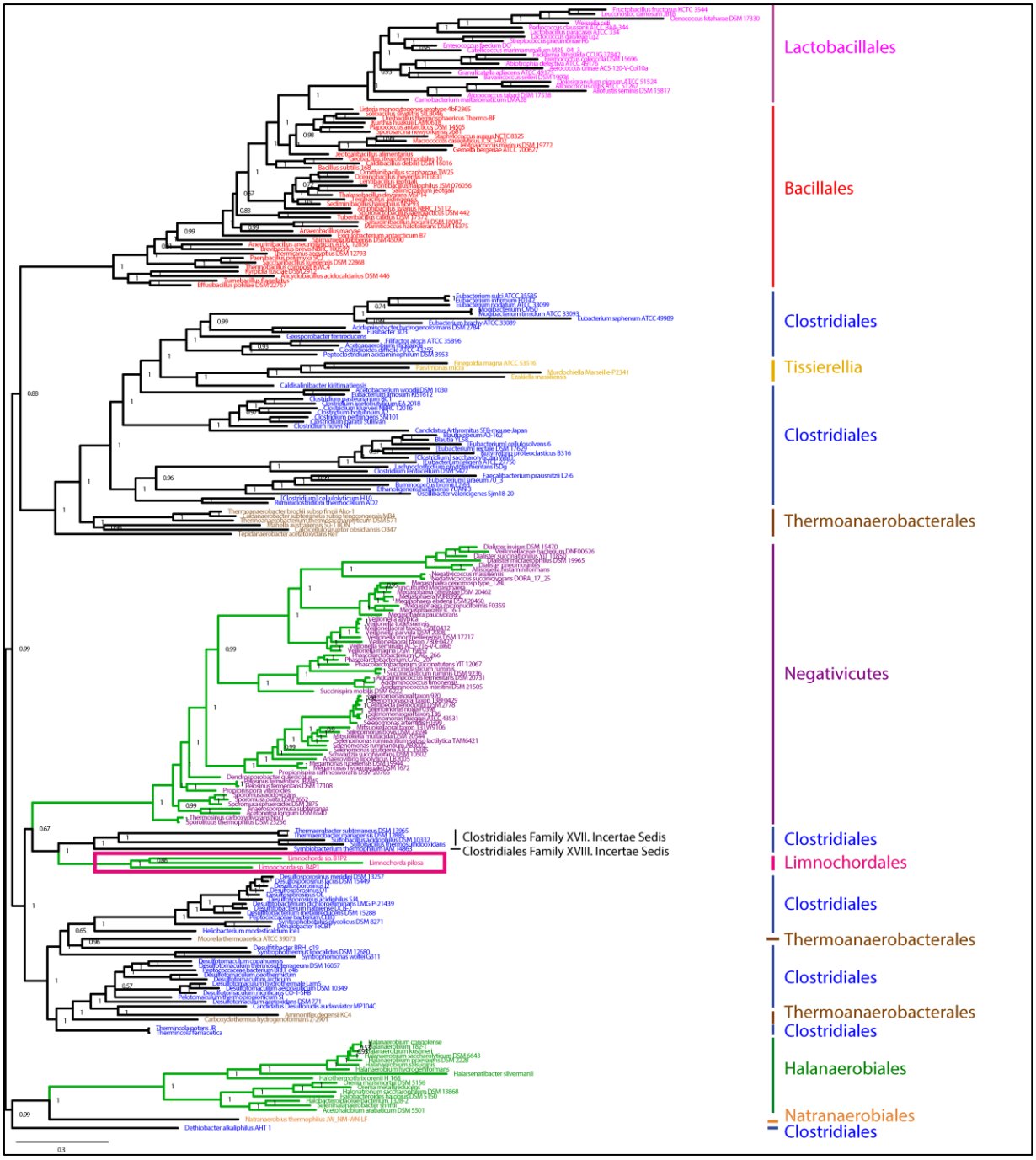

Figure 1. Phylogenetic placement of Limnochordales. Phylum-level phylogeny of the Firmicutes. Bayesian phylogeny of the Firmicutes based on a concatenation of 47 orthologous ribosomal proteins comprising 5,551 amino acid positions. The tree was calculated by PhyloBayes and the CAT+GTR+Γ4 model (Lartillot et al., 2009). Values at nodes represent Bayesian posterior probabilities. The scale bar represents the average number of substitutions per site.

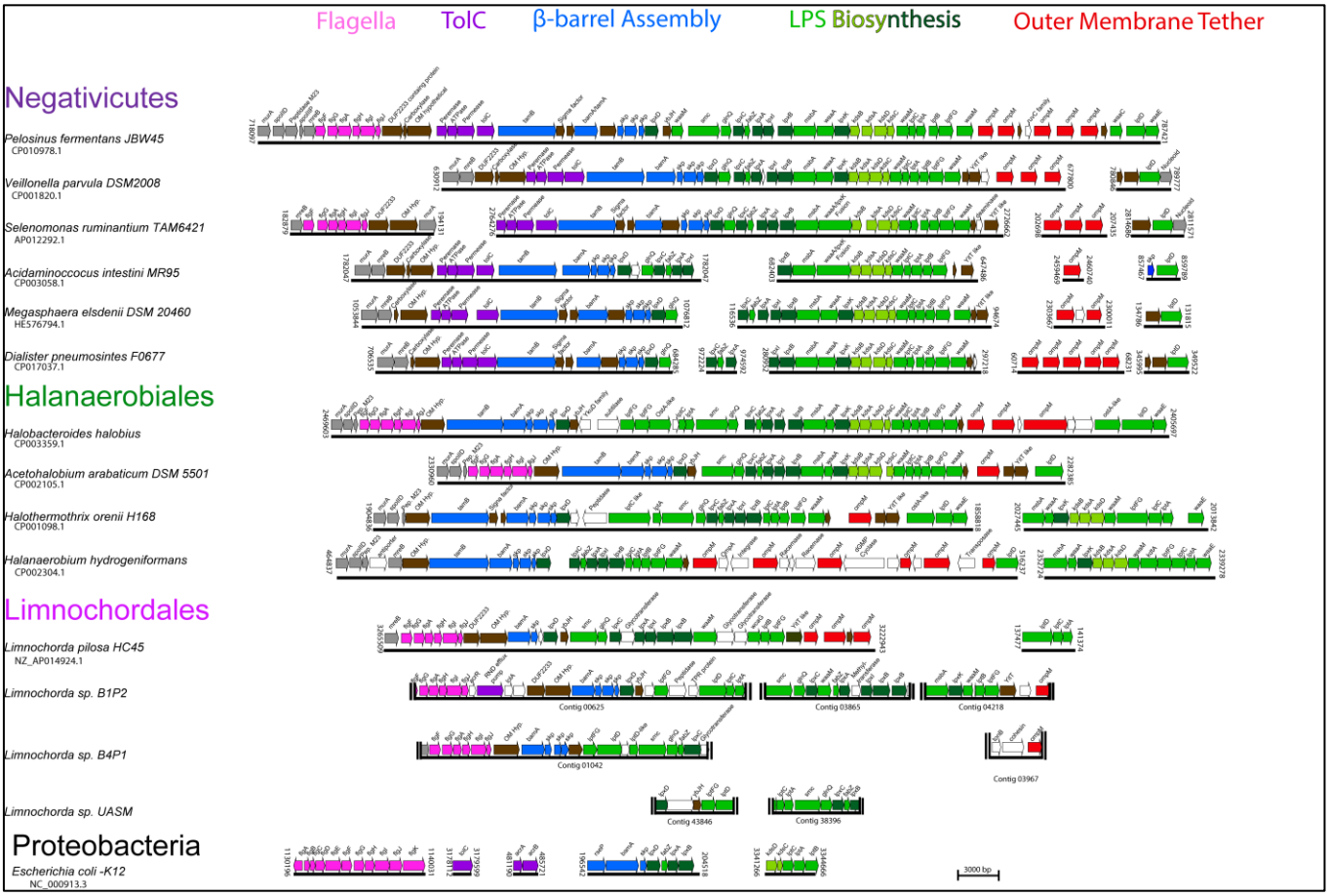

Figure 2. Colocalization of genes encoding LPS synthesis and transport, OMP assembly and structural OMPs in the Negativicutes, Halanaerobiales, Limnochordales, and *Escherichia coli*. One representative for each genus with complete genomes of diderm Firmicutes is shown, with all Limnochordales (for full distribution and accession numbers see Supplementary file 1). Genome accession number is stated under the name while nucleotide position is given at the start and end of each continuous diderm locus (solid black bar). Contigs have their names listed underneath and are bordered by double black bars. Genes are coloured according to their functional class: LPS synthesis and transport (green), OMP assembly (blue), flagellum (light pink), OM-PG attachment (red), Conserved hypothetical (brown), efflux (purple), and conserved peripheral genes (grey) (see text for discussion). White boxes indicate proteins not known to being related to the OM or nonconserved proteins whose connection with the OM is unclear. Genomic synteny was investigated using the interactive web-based visualization tool SyntTax (Oberto, 2013).

Some unique characteristics are evident in the Limnochordales genomes. For example, they lack homologues of tamB, the IM β-barrel assembly chaperone of a putative ancestral TAM/BAM system that has a conserved co-localisation with the OM component *bamA* and the chaperone *skp* in both Negativicutes and Halanaerobiales (Antunes et al., 2016), as well as other bacteria (Heinz et al., 2015). We could only detect *bamA* and *skp* within the Limnochordales cluster. *tamB* homologs are absent in a few deep branching phyla as well, most notably the Thermotoga (Heinz et al., 2015), suggesting that Limnochordales may not be unique in this aspect. *tamB* may be replaced by *asmA*, a homolog of *tamB* with unclear function (Prieto et al., 2009), which is present at a different genetic locus.

Interestingly, the genomic locus of Limnochordales contains two hypothetical OM proteins (DUF2233-containing protein and YitT-like protein (Figure 2) that we had previously identified as being conserved in Negativicutes and Halanaerobiales (Antunes et al., 2016) and were experimentally proven to be in the OM (Antunes et al., 2016; Poppleton et al., 2017). Additionally, we identified two other conserved hypothetical proteins within the gene cluster that we did not previously note: a sigma factor (Sigma 28, FliA family) and *ybjH*, a β-barrel protein with unknown function

(Figure 2). The fact that these proteins are conserved in all three diderm Firmicutes lineages strongly suggests that they may be involved in processes linked to the OM and makes them priority targets for experimental characterization. Finally, Limnochordales appear to have a reduced complement of the genes involved in LPS biosynthesis, which is otherwise present in both Halanaerobiales and Negativicutes. Most notably, they lack *lpxK*, an enzyme which phosphorylates lipid A, *lpxL/M*, responsible for the addition of a lauryl group, as well as KdsA-D, responsible for KDO biosynthesis. These functions may be replaced by hypothesized glycotransferases in the *L. pilosa* genome (Figure 2).

Phylogenetic analysis of LPS core genes is consistent with a vertical inheritance of the OM in the Limnochordales

Limnochordales were included in an updated phylogeny of the core genes of LPS biosynthesis (*lpxABCD*) including representatives from all major LPS producing bacterial phyla (Figure 3). The tree is completely congruent with our previous analysis (Antunes et al., 2016), demonstrating no recent transfer of the LPS biosynthesis machinery among major phyla. Moreover, Limnochordales robustly cluster with Halanaerobiales and Negativicutes. This strengthens our previous hypothesis that the four core LPS genes, and by extension, the other components of the large genomic locus, were present in the ancestor of all Firmicutes and inherited vertically in the three diderm lineages.

Previous analyses on the evolution of KDO-LPS biosynthesis in bacteria have proposed a stepwise acquisition of LPS biosynthesis genes plateauing at Escherichia coli with a full complement of genes, with the Cyanobacteria and Dictyoglomi representing the ancestral four-genes reduced state (Opiyo et al., 2010). As Limnochordales have also a reduced complement of LPS biosynthetic genes and no KDO genes, we mapped the presence/absence of these genes upon our tree (Figure 3, Supplemental table 2). The inclusion of a larger taxonomic sampling and of the three diderm Firmicutes lineages now indicates the presence of reduced pathways in a number of bacterial lineages other than Limnochordales and Dictyoglomi, namely Deinococcus-Thermus, Dictyoglomi, and the Atribacteria (Figure 3). This might mean that this reduction might have arisen multiple times independently during bacterial diversification led to a wide array of LPS molecules. In the Cyanobacteria we found the biosynthetic capabilities to produce KDO, however they did not encode *waaF*, the attachment enzyme. The Cyanobacteria have a diverse array of LPS types, both with and without KDO (Durai et al., 2015), and is not likely to be the ancestral state. As for *L. pilosa*, it is possible that these genes were replaced by non-homologous glycotransferases, a hypothesis that remains to be investigated experimentally. These novel LPS types need isolation and structural characterization, as they might have interesting properties, as some types of Cyanobacterial LPS have benefits to human health (Durai et al., 2015).

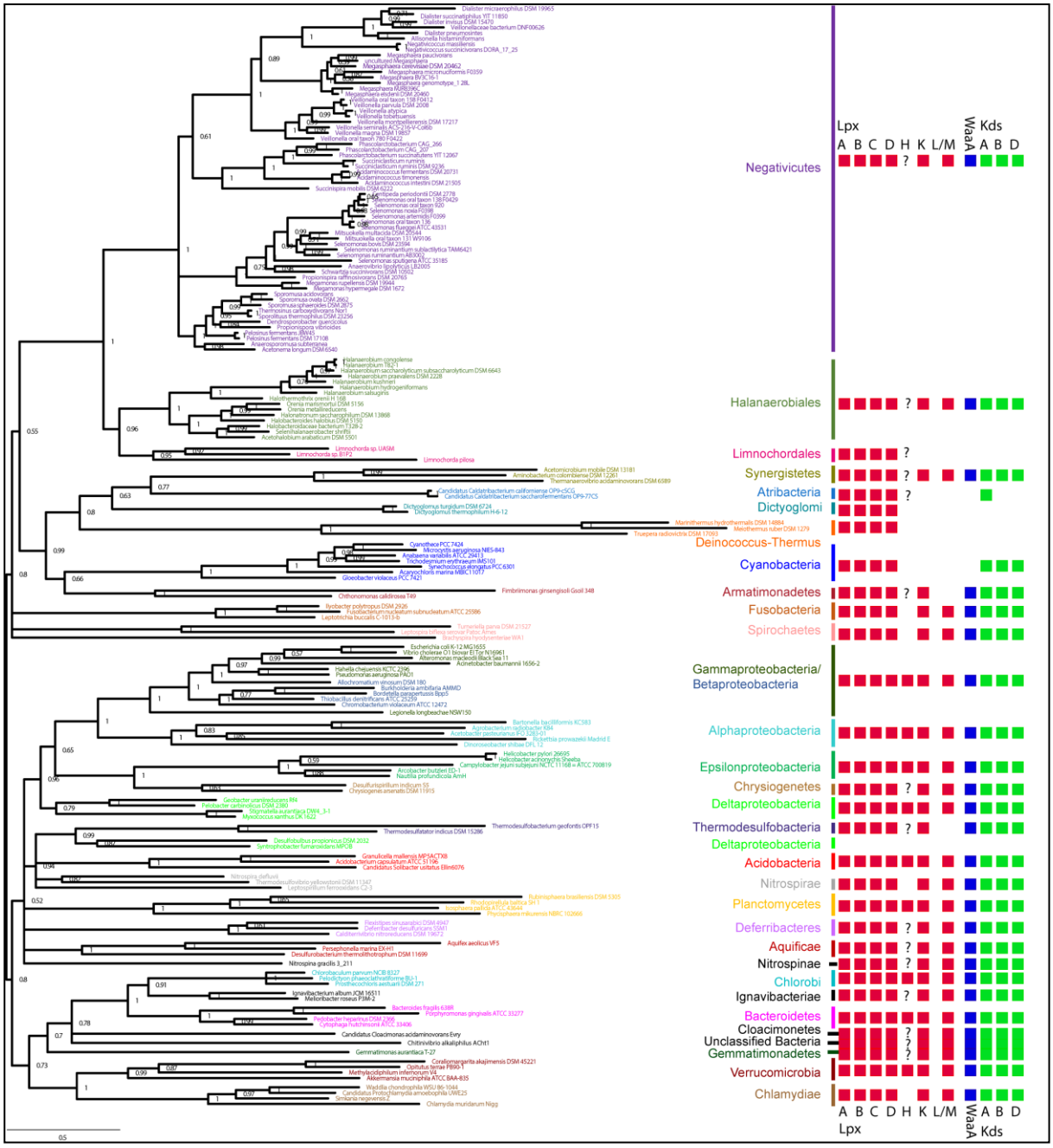

Figure 3. Phylogenetic tree of core LPS components and distribution of KDO-LPS biosynthesis genes. Bayesian phylogeny based on a concatenation of orthologs of the four core components of the LPS biosynthesis pathway (lpxABCD), comprising 898 amino acid positions and the CAT+GTR+Γ4 model. Values at nodes represent Bayesian posterior probabilities. The scale bar represents the average number of substitutions per site. For mapping the distribution of LPS biosynthesis genes, all were queried with HMM-based profiles (Mistry *et al.*, 2013), except for LpxH, for which no specific profile could be obtained and whose distribution was therefore taken from (Opiyo *et al.*, 2010), but was left unknown for the newly analyzed genomes. The full box indicates presence, while question marks indicate not known.

Conclusions

The presence of a third diderm lineage in the Firmicutes confirms and extends our previous analysis and the hypothesis that the ancestor of all Firmicutes possessed an LPS outer membrane (Antunes et al., 2016). In particular, while the possibility could not be totally excluded that the last common ancestor of the Negativicutes transferred the OM to the last common ancestor of the Halanaerobiales, the existence of a third, independent diderm lineage weakens this scenario. It would, in fact, require at least three ancestral transfers, a highly improbable evolutionary event. Moreover, the inclusion of Limnochordales in our analyses leads to inferring even more independent losses of the outer membrane in this phylum than previously thought.

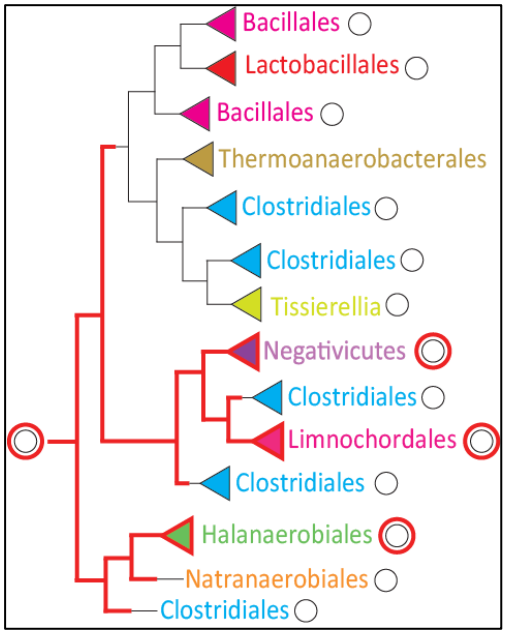

Figure 4. The evolutionary scenario on the Evolution of the OM in the Firmicutes.
The presence of an LPS-OM in the ancestor of all Firmicutes is indicated by a double red circle. Thick red lines represent the vertical inheritance of the ancestral OM in the three diderm lineages (Negativicutes, Halanaerobiales, and Limnochordales), whereas thin black lines represent the multiple OM losses that gave rise to the monoderm cell envelope architecture of most present-day Firmicutes.

In this respect, the study of diderm Firmicutes and their closely related monoderm lineages is key understand the processes that led to OM loss, and the impact on key systems related to the biogenesis and functioning of the cell envelope. Interestingly, the Symbiobacteriaceae, the monoderm clade that is sister to Limnochordales, displays interesting characteristics. One of their members, *Symbiobacterium thermophilum*, was in fact originally thought to be diderm due to its relatively thin PG and an apparent OM in electron microscopy images (Hirata et al., 2000). Indeed, it was later found that this is, in fact, an S-layer, and this was confirmed by the fact that no sequenced genomes of members of the Symbiobacterium clade possess any homologs of the genes coding for OM markers (Ueda et al., 2004). Nonetheless, the cell envelope characteristics and the close phylogenetic relationships of Symbiobacterium to Limnochordales represent a narrower model to study how the OM was lost in most Firmicutes. A provocative yet possible hypothesis is that the apparently reduced number of genes coding for OM components within the Limnochordales may be an indication that this clade is in the process of losing its OM.

Characterization of the Limnochordales cell envelope will add important insights into the nature of the diderm Firmicutes outer membrane and further highlight the diderm/monoderm transition in Bacteria. Finally, we anticipate that additional Firmicutes diderm lineages may be found through genomic and microscopy exploration of the vast diversity of this major bacterial phylum.

Materials and Methods

Phylogenetic analyses

All methods were taken in their entirety from our previous work (Antunes et al., 2016), with the exception of sequence alignment which was performed with MAFFT v7.055 (Yamada et al., 2016).

Metagenome Discovery and Assembly

A database consisting of all metagenomes available on JGI IMG (Chen et al., 2017) was constructed. This metagenomics database was translated using Prodigal (Hyatt et al., 2010) with default settings for metagenomes. A subdatabank was constructed that consisted of only the proteins that elucidated a positive match to the HMM for RNA polymerase β' (PF00623). This databank was queried using the for RNA polymerase β' of L. pilosa (BAS29096.1) by BLAST (Camacho et al., 2009). The top hits were further analyzed and genome reconstruction was attempted. To reconstruct the genomes tetramer frequency was calculated (Dick et al., 2009) and the genomes were binned with emergent self-organizing map (ESOM) (Ultsch and Mörchen, 2005). These genomes were checked for completeness, heterogeneity, and contamination with CheckM (Parks et al., 2015).

References

Alauzet, C., Marchandin, H., Courtin, P., Mory, F., Lemée, L., Pons, J.-L., *et al.* (2014) Multilocus analysis reveals diversity in the genus Tissierella: Description of Tissierella carlieri sp. nov. in the new class Tissierellia classis nov. *Syst Appl Microbiol* **37**: 23–34

Antunes, L.C., Poppleton, D., Klingl, A., Dupuy, B., Brochier-Armanetd, C., Beloin, C., and Gribaldo, S. (2016) Phylogenomic analysis supports the ancestral presence of LPS outer membranes in the Firmicutes. *Elife* e14589

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421

Chen, I.-M.A., Markowitz, V.M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., *et al.* (2017) IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res* **45**: D507–D516

Dick, G.J., Andersson, A.F., Baker, B.J., Simmons, S.L., Thomas, B.C., Yelton, A.P., and Banfield, J.F. (2009) Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10**: R85

Durai, P., Batool, M., and Choi, S. (2015) Structure and Effects of Cyanobacterial Lipopolysaccharides. *Mar Drugs* **13**: 4217–30 http://www.ncbi.nlm.nih.gov/pubmed/26198237. Accessed July 31, 2017.

Fukui, M., Kojima, H., and Watanabe, M. (2016) Complete genome sequence and cell structure of Limnochorda pilosa, a Gram-negative spore-former within the phylum Firmicutes. *Int J Syst Evol Microbiol* **66**: 1330–1339

Heinz, E., Selkrig, J., Belousoff, M., and Lithgow, T. (2015) Evolution of the Translocation and Assembly Module (TAM). *Genome Biol Evol* evv097.

Hirata, A., Ueda, K., Saitoh, Y., Park, M.J., Yamashita, N., Nishida, H., *et al.* (2000) Symbiobacterium thermophilum gen. nov., sp. nov., a symbiotic thermophile that depends on co-culture with a Bacillus strain for growth. *Int J Syst Evol Microbiol* **50**: 1829–1832 h

Lartillot, N., Lepage, T., and Blanquart, S. (2009) PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**: 2286–2288.

Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., and Punta, M. (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* **41**: e121

Oberto, J. (2013) SyntTax: a web server linking synteny to prokaryotic taxonomy. *BMC Bioinformatics* **14**: 4

Opiyo, S.O., Pardy, R.L., Moriyama, H., and Moriyama, E.N. (2010) Evolution of the Kdo2-lipid A biosynthesis in bacteria. *BMC Evol Biol* **10**:

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–55

Poppleton, D.I., Duchateau, M., Hourdel, V., Matondo, M., Flechsler, J., Klingl, A., *et al.* (2017) Outer Membrane Proteome of Veillonella parvula: A Diderm Firmicute of the Human Microbiome. *Front Microbiol* **8**: 1215

Prieto, A.I., Hernandez, S.B., Cota, I., Pucciarelli, M.G., Orlov, Y., Ramos-Morales, F., *et al.* (2009) Roles of the Outer Membrane Protein AsmA of Salmonella enterica in the Control of marRAB Expression and Invasion of Epithelial Cells. *J Bacteriol* **191**: 3615–3622

Ueda, K., Yamashita, A., Ishikawa, J., Shimada, M., Watsuji, T., Morimura, K., *et al.* (2004) Genome sequence of Symbiobacterium thermophilum, an uncultivable bacterium that depends on microbial commensalism. *Nucleic Acids Res* **32**: 4937–4944 http://www.ncbi.nlm.nih.gov/pubmed/15383646. Accessed August 11, 2017.

Ultsch, A., and Mörchen, F. (2005) ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM.

Watanabe, M., Kojima, H., and Fukui, M. (2015) Limnochorda pilosa gen. nov., sp. nov., a moderately thermophilic, facultatively anaerobic, pleomorphic bacterium and proposal of Limnochordaceae fam. nov., Limnochordales ord. nov. and Limnochordia classis nov. in the phylum Firmicutes. *Int J Syst Evol Microbiol* **65**: 2378–2384

Whitman, W.B. (2011) *Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*. 2nd Editio., Springer, .

2 Describe the outer membrane of diderm Firmicutes

Although the work on phylogeny was extremely informative, further characterization of the diderm Firmicute envelope was necessary to support and enhance my previous work. In this section, I present two articles: The first is the OM proteome of *V. parvula* with subsequent analysis, while the second is the in-depth phylogenomic analysis of a single mechanism within all major Bacterial groups.

2.1 Outer Membrane Proteome of *Veillonella parvula*

Phylogeny and computation can only progress research so far and it must be built on fundamental bench science. In this work, I obtained the OM proteome of *V. parvula*. I chose *V. parvula* for multiple reasons, the first and not the least is that the Qi laboratory in the United States had been working on a genetic system that seemed promising for further analysis of the species; there was no genetic system in any diderm Firmicute to this point. Another reason is that *V. parvula* is closely tied to human health, both positively and negatively, and our results may enable further research on different components of *Veillonella's* OM components.

These results were a true merging of bioinformatics and wet experimental work that were combined for some of the best work of this thesis. As we were unable to perform any mutational assay as a control for our tandem mass spectrometry I combined multiple prediction programs to obtain a suitable substitute.

In this work, we were able to confirm all of the assertions for the presence of OM components within the diderm Firmicutes. Additionally, we were able to identify components that were undescribed in the eLIFE paper, such as LptD the OM LPS translocase.

Surprisingly, we found many unexpected features within the genome and proteome. A large degree of clustering was found for all the OM genes, with the adhesins forming one cluster, the TonB dependent receptors forming another, and the previously described diderm cluster being a third.

These results may be a springboard for other work on *Veillonella.* The detailed computational analysis combined with the mass spectrometry opened more questions than it answered. For instance is the dichotomy of the Firmicute Flagella/adhesins a regular adaptation? What function do the expressed conserved hypothetical proteins do within these cells? What about the conserved domains? With such an important commensal these questions are necessary and need answers. The most abundant protein by mass-spectrometry sprang me on to the article after this on OM attachment.

We recently published this work in Frontiers in Microbiology

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 June 2017
doi: 10.3389/fmicb.2017.01215

Outer Membrane Proteome of *Veillonella parvula:* A Diderm Firmicute of the Human Microbiome

Daniel I. Poppleton[1], Magalie Duchateau[2], Véronique Hourdel[2], Mariette Matondo[2], Jennifer Flechsler[3], Andreas Klingl[3], Christophe Beloin[4] and Simonetta Gribaldo[1]**

[1] Unité de Biologie Moléculaire du Gène chez les Extrêmophiles, Département de Microbiologie, Institut Pasteur, Paris, France, [2] Unité de Spectrométrie de Masse Structurale et Protéomique, Plateforme Protéomique, Département de Biologie Structurale et Chime, Institut Pasteur, USR 2000 Centre National de la Recherche Scientifique, Paris, France, [3] Pflanzliche Entwicklungsbiologie und Elektronenmikroskopie, Department I. Botanik, Biozentrum der LMU München, Planegg-Martinsried, Germany, [4] Unité de Génétique des Biofilms, Département de Microbiologie, Institut Pasteur, Paris, France

OPEN ACCESS

Edited by:
Mickael Desvaux, Institut National de la Recherche Agronomique (INRA), France

Reviewed by:
Iain Sutcliffe, Northumbria University, United Kingdom
Jean Armengaud, Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), France

***Correspondence:**
Christophe Beloin
christophe.beloin@pasteur.fr
Simonetta Gribaldo
simonetta.gribaldo@pasteur.fr

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *17 March 2017*
Accepted: *14 June 2017*
Published: *30 June 2017*

Citation:
Poppleton DI, Duchateau M, Hourdel V, Matondo M, Flechsler J, Klingl A, Beloin C and Gribaldo S (2017) Outer Membrane Proteome of Veillonella parvula: A Diderm Firmicute of the Human Microbiome. Front. Microbiol. 8:1215. doi: 10.3389/fmicb.2017.01215

Veillonella parvula is a biofilm-forming commensal found in the lungs, vagina, mouth, and gastro-intestinal tract of humans, yet it may develop into an opportunistic pathogen. Furthermore, the presence of *Veillonella* has been associated with the development of a healthy immune system in infants. *Veillonella* belongs to the Negativicutes, a diverse clade of bacteria that represent an evolutionary enigma: they phylogenetically belong to Gram-positive (monoderm) Firmicutes yet maintain an outer membrane (OM) with lipopolysaccharide similar to classic Gram-negative (diderm) bacteria. The OMs of Negativicutes have unique characteristics including the replacement of Braun's lipoprotein by OmpM for tethering the OM to the peptidoglycan. Through phylogenomic analysis, we have recently provided bioinformatic annotation of the Negativicutes diderm cell envelope. We showed that it is a unique type of envelope that was present in the ancestor of present-day Firmicutes and lost multiple times independently in this phylum, giving rise to the monoderm architecture; however, little experimental data is presently available for any Negativicutes cell envelope. Here, we performed the first experimental proteomic characterization of the cell envelope of a diderm Firmicute, producing an OM proteome of *V. parvula*. We initially conducted a thorough bioinformatics analysis of all 1,844 predicted proteins from *V. parvula* DSM 2008's genome using 12 different localization prediction programs. These results were complemented by protein extraction with surface exposed (SE) protein tags and by subcellular fractionation, both of which were analyzed by liquid chromatography tandem mass spectrometry. The merging of proteomics and bioinformatics results allowed identification of 78 OM proteins. These include a number of receptors for TonB-dependent transport, the main component of the BAM system for OM protein biogenesis (BamA), the Lpt system component LptD, which is responsible for insertion of LPS into the OM, and several copies of the major OmpM protein. The annotation of *V. parvula's* OM proteome markedly extends previous inferences on the nature of the cell envelope of Negativicutes, including the experimental evidence of a BAM/TAM system for OM protein biogenesis and of a complete Lpt system

for LPS transport to the OM. It also provides important information on the role of OM components in the lifestyle of *Veillonella*, such as a possible gene cluster for O-antigen synthesis and a large number of adhesins. Finally, many OM hypothetical proteins were identified, which are priority targets for further characterization.

Keywords: Negativicutes, proteomics, BAM/TAM complex, adhesins, OmpM, LPS, Lpt system, O-antigen

INTRODUCTION

Veillonella parvula, an anaerobic coccus, was discovered and described 120 years ago when Veillon and Zuber isolated it from an appendicitis abscess (Veillon and Zuber, 1898). Sixty years later the same microbe was used in the first observation of an outer membrane (OM) (Bladen and Mergenhagen, 1964) thereby demonstrating a key difference between the Gram-negative and Gram-positive cell envelope architecture. Since then, 13 other *Veillonella* species have been isolated and described from both humans and rodents (Euzeby, 1997). *Veillonella* strains are found in several niches of the human body including the mouth (Do et al., 2015), lungs, gastrointestinal tract (Rosen et al., 2014), and vagina (Africa et al., 2014). *V. parvula* may serve an important role in many of these environments, however its dominant niche is as a secondary colonizer in the mouth (Griffen et al., 2012). This normal component of the human microbiome may be an opportunistic pathogen; its presence has been associated with several disease states and was the primary infectious agent in at least fifty cases since the 1970's, both with and without an underlying condition (Hirai et al., 2016). In direct opposition to its role as an opportunistic pathogen, recent data suggests that *Veillonella* may perform a protective role and aid in early childhood immune system development. Epidemiological studies of infants have demonstrated that presence of *Veillonella* is negatively correlated with asthma (Arrieta et al., 2015), bronchiolitis (Hasegawa et al., 2016), and autism (Strati et al., 2017).

In addition to being an important component of the human microbiome, *Veillonella,* and other members of the Negativicutes are evolutionary enigmas. They phylogenetically belong to the Firmicutes (low GC Gram-positives or monoderms), yet possess an OM with lipopolysaccharides (LPS) similar to Gram-negative (diderm) bacteria (Zeikus et al., 1983; Tocheva et al., 2011; Campbell et al., 2014). We recently showed that the genomes of Negativicutes contain genes encoding typical diderm OM machinery including flagellar rings typical of diderms, Type 4 pili (T4P) secretin, and an ancestral form of the BAM/TAM system for OM β-barrel assembly (Antunes et al., 2016). Furthermore, most OM-related genes were found to reside in a single genomic cluster (Antunes et al., 2016), hereafter referred to as the "diderm cluster," that included genes responsible for LPS biosynthesis. Through phylogenomic analysis, we proposed that the diderm cell envelope of Negativicutes represents an ancestral characteristic that was present in the ancestor of all Firmicutes, and was lost multiple times independently to give rise to the monoderm cell envelope in this phylum (Antunes et al., 2016). Although this work presented important perspectives on the OM of *Veillonella,* it was performed *in silico* and cannot tell us which proteins are true components of the OM and whether they are expressed. These predictions can be validated using experimental proteomic approaches, such as those performed on *Bacteroides fragilis* (Wilson et al., 2015) and *Actinobacillus pleuropneumoniae* (Chung et al., 2007).

Only two outer membrane proteins from the Negativicutes have been cloned and studied: OmpM and hemagglutinin-like adhesins. OmpM is an alternative method for tethering the OM to the peptidoglycan by binding of polyamine-modified peptidoglycan through an S-Layer homology (SLH) domain and a transmembrane β-barrel for OM attachment (Kojima and Kamio, 2012). This contrasts with *Escherichia coli's* Braun's lipoprotein (Lpp), which covalently binds PG and integrates into the OM via a lipid moiety (Braun and Rehn, 1969). In the case of adhesins, eight were found in *Veillonella atypica*; one of which, Hag1, was shown to bind human buccal cells and oral streptococci (Zhou et al., 2015).

In order to extend these bioinformatic and experimental data, we have performed the first proteomic analysis of the OM of *V. parvula*. By combining bioinformatic prediction and subcellular fractionation, we have obtained fundamental insight into the nature of these unique cell envelopes and the lifestyle of *Veillonella,* which will help future work on this important component of the human microbiome.

MATERIALS AND METHODS

Bioinformatic Prediction

To perform *in silico* prediction, we used 12 distinct bioinformatic programs on all 1,844 proteins encoded in the *V. parvula* DSM 2008 genome. Initial prediction was performed using three general prediction programs for IM (Inner Membrane), cytoplasmic, periplasmic, secreted, and OM: PSORTb 3.0.2 (Yu et al., 2010) using default input parameters for Gram-negative bacteria, CELLO 2.5 (Yu et al., 2004) with default input parameters for Gram-negative bacteria, and SOSUI-GramN (Imai et al., 2008) with default parameters. These analyses were complemented with prediction of transmembrane helices by TMHMM 2.0 (Krogh et al., 2001). β-Barrels were predicted using BOMP (Berven et al., 2004) with the additional BLAST option. LipoP 1.0 (Juncker et al., 2003) was used to refine cytoplasmic and IM prediction. Positive lipoprotein prediction was defined as a consensus from PRED-LIPO (Bagos et al., 2008) and LipoP. TAT secreted proteins were identified as a consensus from PRED-TAT (Bagos et al., 2010) and TatP (Bendtsen et al., 2005). Positive SEC signal sequences were defined as a two out of three or greater concurrent result of SignalP, PRED-LIPO, and Phobius (Kall et al., 2007). Extended signal peptides (ESP) were queried using HMMER (Mistry et al., 2013) with the ESPR

domain (PF13018) from PFAM within an E-value cutoff of 0.001.

Positive prediction for cytoplasmic proteins was determined by a three out of four or greater concurrent results of PSORT, CELLO, LipoP, and SOSUI. IM prediction was determined by three out of five or greater concurrent results of PSORT, CELLO, TMHMM, LipoP, and SOSUI. Positive OM prediction was determined by three out of four or greater concurrent results of PSORT, CELLO, BOMP, and SOSUI. Positive periplasmic prediction and secreted proteins were determined by a two out of three or greater concurrent results of PSORT, CELLO, and SOSUI.

Search for Conserved Domains and Homologs

To identify homologs of experimentally identified proteins we ran a BLAST 2.2.30+ (Camacho et al., 2009) search against two local databases. The first contains a selection of 255 representative Firmicutes and the second contains all 112 currently available Negativicutes genomes. Default settings were used except for an E-value cutoff of 0.0001. Protein domains were searched using the conserved domain database (CDD) from NCBI (Marchler-Bauer et al., 2014), PFAM (Finn et al., 2015), InterPro (Finn et al., 2017), and PANTHER 10.0 (Mi et al., 2015) with the required E-value cutoff of 0.00001. Protein folds were predicted with Phyre2 (Kelley et al., 2015).

Outer Membrane Protein Extraction

The extraction protocol was modified from Thein et al. (2010). Fifty milliliters of *V. parvula* DSM 2008 were grown anaerobically in triplicate to an optical density at 600 nm of 0.4 (10 h after a 1:100 dilution) in BHIL (BHI (Bacto) + 0.6% sodium L-lactate). The bacteria were harvested by centrifugation at 15,000 × g for 20 min at 4°C and resuspended in an equal volume of Tris-HCl (0.1 M pH 7.3 supplemented with 7 mg of DNAse). Cells were washed two additional times and suspended in 6 ml of the same buffer. Cells were lysed by French press at 10,000 kPa for four passes in Tris-HCl (20 mM pH 7.3). Cellular debris were pelleted by centrifugation at 15,000 × g for 20 min at 4°C. The supernatant was collected and the pellet was discarded. This step was repeated an additional time. Supernatant was then diluted with ice cold 0.1 M Na_2CO_3 pH 11 to a volume of 60 ml and stirred for 1 h at 4°C. The suspension was separated at 120,000 × g for 1 h at 4°C and the pellet washed in an equal volume of Tris-HCl (0.1 M pH 7.3) and spun at 85,000 × g for 20 min at 4°C. The wash was repeated twice. The pellet was resuspended in ddH_2O and proteins were TCA precipitated before MS analysis.

Surface Exposed and Control Sample Extraction

For extraction of surface exposed (SE) fraction, as well as for control samples, a protocol was modified from Voss et al. (2014). Bacteria were grown in triplicate and harvested as in the OM protein extraction. Five milliliters of bacterial cells were washed three times in PBS (0.1 M PO_4, 0.15 M NaCl pH 8.0). Cells were resuspended in PBS containing 20 mM of NHS-PEG4-Biotin (Thermo Fisher Scientific) and incubated on ice for 30 min. Reaction was quenched by washing cells in quench buffer (PBS + 100 mM Glycine) three times. Control samples (WC) were treated identically except that they were incubated in PBS without labeling substrate. Cells were lysed by French press at 10,000 kPa for four passes in radioimmunoprecipitation (RIPA) buffer [25 mM Tris-HCl, pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, and 0.1% sodium dodecyl sulfate (SDS) containing a 1:100 dilution of protease inhibitors] and cellular debris were removed by centrifugation at 15,000 × g for 20 min at 4°C twice. Control sample was then TCA precipitated. For SE fraction 1 ml high Capacity Streptavidin Resin columns (Thermo Fisher Scientific) and reagents were equilibrated to room temperature. PBS and elution buffer (8 M guanidine-HCl, pH 1.5) were filtered and degassed. Column was equilibrated with five column volumes of PBS at a flow rate of 0.2 mL/min on an AKTA-FPLC. Sample was then applied to the column using the same rate and washed with 10 column volumes of PBS before elution in a single column volume of elution buffer. Eluate then underwent MS analysis.

Trichloroacetic Acid (TCA) Precipitation

TCA was added to OM samples and WC samples to a final concentration of 20%. The precipitate was spun at 6,500 × g for 1 h at 4°C. The pellet was washed with 800 µl of −20°C acetone overnight and then spun at 6,500 g for 1 h at 4°C. The washing procedure was performed two additional times before resuspension or storage at −20°C.

LPS Extraction and Visualization

E. coli strains were kindly provided by Laurent Debarbieux. *E. coli* (81009, 81009Δ*waaF*, E47a, E47aΔ*rfb3*; Szijarto et al., 2014) and *V. parvula* DSM 2008 were grown in triplicate to an OD of 0.4. LPS was isolated using hot phenol extraction without modification as previously described (Davis et al., 2012). LPS was then resolved onto either a 0.1% SDS 13% or 17% PAGE gel and visualized with Pro-Q Emerald 300 (ThermoFisher Scientific) as per the manufacturer's instructions.

In-Gel Protein Digestion

Protein samples were loaded on a 0.1% SDS 12% PAGE gel. After the electrophoretic migration, the gel was stained with Coomassie Blue, each band of interest was cut, and in-gel tryptic digestion was performed as described previously (Wilm et al., 1996). Briefly, gel slices were washed in 100 mM ammonium bicarbonate for 15 min, followed by several washing steps to eliminate the stain in 100 mM ammonium bicarbonate/acetonitrile (1:1). Samples were reduced (10 mM DTT in 100 mM ammonium bicarbonate, 30 min at 56°C) and alkylated (55 mM iodoacetamide in 100 mM ammonium bicarbonate, 30 min at room temperature in the dark). Proteins were digested by 250 ng Sequencing Grade Modified Trypsin (Promega, Madison, WI, USA) in 10 mM ammonium bicarbonate overnight at 37°C. Resulting peptides were extracted, dried in the Speed-Vac and then resuspended in water/acetronitrile/formic acid (98:2:0.1).

Liquid Protein Digestion

Tryptic digestion was performed by eFASP (enhanced Filter-Aided Sample Preparation) as described previously (Erde et al., 2014). Samples were buffer-exchanged into buffer containing 8 M urea, 100 mM ammonium bicarbonate, and 0.2% deoxycholic acid using Tween passivated 30-kDa cut-off Amicon filters (Millipore Inc., USA). Samples were reduced with 5 mM tris (2-carboxyethyl) phosphine (TCEP) for 30 min at room temperature, alkylated in iodoacetamide (50 mM final, 30 min room temperature in the dark), treated with 1 μg of sequencing-grade modified trypsin (Promega, USA) overnight at 37°C under agitation. Peptides were recovered by centrifugation.

Mass Spectrometry Analysis

Trypsin-digested peptides coming from gel samples were analyzed by nano LC-MS/MS using an Ultimate 3000 system (Dionex, Amsterdam, The Netherlands) coupled to an LTQ-Orbitrap Velos. File microliters of each sample were loaded on a C_{18} pre-column (300 μm inner diameter × 5 mm; Dionex) at 30 μL/min in 2% ACN, 0.1% FA. After 5 min of desalting, the pre-column was switched online an in-house packed 15 cm nano-HPLC column (75 μm inner diameter) with C_{18} resin (3 μm particles, 100 Å pore size, ReproSil-Pur Basic C_{18}, Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) and equilibrated in 95% solvent A (2% ACN, 0.1% FA) and 5% solvent B (80% ACN, 0.08% FA). The peptides were eluted using a 2–50% gradient of solvent B during 60 min at 300 nL/min flow rate. The LTQ-Orbitrap Velos (Thermo Fisher Scientific, Bremen) was operated in data-dependent acquisition mode with the XCalibur software (Thermo Fisher Scientific, Bremen). Survey scan MS were acquired in the Orbitrap on the 300–2,000 m/z range with the resolution set to a value of 60,000 at $m/z = 400$. The 10 most intense ions per survey scan were selected for CID fragmentation, and the resulting fragments were analyzed in the linear trap (LTQ). The dynamic exclusion was enabled with the following settings: repeat count, 1; repeat duration, 30 s; exclusion list size, 500; and exclusion duration, 20 s.

Digests coming from liquid samples were analyzed on an Orbitrap Q Exactive Plus mass spectrometer (Thermo Fisher Scientific, Bremen) coupled with an EASY nLC 1000 chromatography system (Thermo Fisher Scientific). Sample was loaded on an in-house packed 40 cm nano-HPLC column (75 μm inner diameter) with C_{18} resin (1.9 μm particles, 100 Å pore size, Reprosil-Pur Basic C_{18}-HD resin, Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) and equilibrated in 98% solvent A (H_2O, 0.1% FA) and 2% solvent B (ACN, 0.1% FA). Peptides were eluted using a 2–45% gradient of solvent B during 120 or 240 min at 250 nL/min flow rate. The instrument method for the Q Exactive Plus was set up in the data dependent acquisition mode using XCalibur 2.2 software (Thermo Fisher Scientific, Bremen). After a survey scan in the Orbitrap (resolution 70,000 at m/z 400), the 10 most intense precursor ions were selected for higher-energy collision dissociation (HCD) fragmentation with a normalized collision energy set to 28. Charge state screening was enabled, and precursors with unknown charge state or a charge state of 1 and >7 were excluded. Dynamic exclusion was enabled for 35 s.

In order to increase throughput and sensitivity methods used to acquire the data with these two mass spectrometer have been optimized in accordance with recommended settings from Thermo Fisher Scientific using HeLa extract sample (from Thermo) and an in-house complex quality control sample.

Data Processing and Analysis

Raw data were searched using MaxQuant (version 1.4.1.2) (Cox and Mann, 2008; Cox et al., 2014; with the Andromeda search engine) against *V. parvula* Uniprot database (1,844 entries, version 10-2015) and concatenated with known MS contaminants and reversed sequences of all entries. Andromeda searches were performed choosing trypsin as specific enzyme with a maximum number of two missed cleavages. Possible modifications included carbamidomethylation (Cys, fixed), oxidation (Met, variable), and Nter acetylation (variable). Maximum peptide charge was set to seven and five amino acids were required as minimum peptide length. Most peptides (~89%) were identified with charge states $< +3$. Less than 10% of peptides were identified with a charge state $> +4$. As you may see on the graph, peptides (~0.1%) were identified with a charge state of +6 (for details see PSM charges Pourcentage file deposited in PRIDE). Andromeda default settings were used; an initial search was performed using a mass tolerance of 20 ppm, followed by mass recalibration and a main search with a mass tolerance of 5 ppm for parent ions (Tyanova et al., 2016a). In the MS/MS search, mass tolerance was set to 10 ppm for the Q Exactive Plus and to 0.5 Da for the LTQ-Orbitrap Velos. Additional peptides were identified by the "match between run" option with a maximal retention time window of 1 min. One unique peptide to the protein group was required for the protein identification. Only unique peptides were used to distinguish isoforms. If no unique peptide was identified, isoforms were combined into the same "protein group" and the quantification was done for this protein group. A false discovery rate (FDR) cutoff of 1% was applied at the peptide and protein levels. MaxLFQ, Maxquant's label-free quantification (LFQ) algorithm was used to calculate protein intensity profiles across samples (Cox et al., 2014). Data were filtered by requiring a minimum peptide ratio count of two in MaxLFQ. Absolute protein amounts were calculated in Maxquant as the sum of all peptide peak intensities divided by the number of theoretically observable tryptic peptides (intensity-based absolute quantification or iBAQ (Schwanhäusser et al., 2011).

For statistical and bioinformatics analysis, as well as for visualization, Perseus environment was used, which is part of Maxquant (Tyanova et al., 2016b).

The "proteinGroup.txt" file generated with Maxquant was used by Perseus to identify proteins enriched in Biotin samples. Protein identifications were filtered, removing hits to the reverse decoy database as well as proteins only identified by modified peptides or considered as potential contaminant. Protein LFQ intensities were logarithmized. Two valid values out of three were required for each protein for the confident quantification across all replicates and missing values imputed by values simulating low abundance values close to the noise level. For pairwise comparison and identification of enriched proteins, a modified

t-test were applied with permutation-based FDR statistics set to 1% and a S_0 of 1 (Tusher et al., 2001).

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2016) partner repository with the dataset identifier PXD005929.

Transmission Electron Microscopy

A densely grown culture of *V. parvula* DSM 2008 was used for the preparation for transmission electron microscopy (TEM). After concentration of the cells via centrifugation, the cells were either chemically fixed with glutaraldehyde or cryo-fixed with a high-pressure freezer and subsequent freeze substitution.

For chemical fixation, the cell pellet was resuspended in a fixation buffer consisting of 100 mM cacodylate including 2 mM $MgCl_2$, 2.5% glutaraldehyde (final concentration; 25% stock solution), and 270 mM NaCl to adjust the osmolarity of the buffer to the original growth medium. After fixation, the cells were washed five times with buffer, post-fixed for 30 min in 1% osmium tetroxide, washed again two times with buffer and three times with double distilled water. Afterwards, the cells were dehydrated in a graded acetone series, infiltrated with Spurr's resin and polymerized for 72 h at 63°C. The embedded samples were then ultrathin sectioned (50 nm sections), post-stained with lead citrate and visualized in a TEM.

High-pressure freezing was performed with 2 µl of the resuspended concentrated cells in the respective aluminum platelets using a Leica HPM 100 high-pressure freezer (Leica Microsystems GmbH, Wetzlar, Germany). Freeze substitution was performed in a Leica AFS 2 (Leica Microsystems GmbH, Wetzlar, Germany) according to the following protocol (substitution solution): −90°C for 20 h (2% OsO4/acetone), heating to −60°C within 3 h (2% OsO4/acetone), −60°C for 4 h (2% OsO4/acetone), −60°C for 4 h (2% OsO4/0. 2% uranyl acetate/acetone), heating to −30°C within 3 h (2% OsO4/0.2% uranyl acetate/acetone), −30°C for 8 h (2% OsO4/0.2% uranyl acetate/acetone), heating to 0°C (2% OsO4/0.2% uranyl acetate/acetone). After washing three times with ice cold acetone, the samples were infiltrated with epoxy resin (Spurr's resin) and polymerized for 72 h at 63°C. The following steps were identical to the chemical fixation protocol.

Transmission electron microscopy was performed on a Zeiss EM 912 (Carl Zeiss AG, Oberkochen, Germany), operated at 80 kV in the zero-loss mode and equipped with a 2 k × 2 k dual-speed CCD camera (Tröndle Restlichtverstärkersysteme, Moorenweis, Germany).

RESULTS AND DISCUSSION

Bioinformatic Localization Prediction

We started our study with a comprehensive *in silico* analysis of all 1,844 annotated protein-coding genes within the genome of *V. parvula* DSM 2008 (see Section Materials and Methods). The localization prediction of these 1,844 proteins are shown in **Figure 1A**. We used three general prediction programs (PSORT, CELLO, and SOSUI) for inner membrane (IM), cytoplasmic, periplasmic, secreted, and OM, and specific programs for OM (BOMP), IM (TMHMM and LipoP), and cytoplasmic (LipoP) localization. With this strategy, we managed to robustly predict the localization of 85% (1559/1844) of the proteins: 63% (1,171) cytoplasmic, 17% (309) IM, 2% (44) periplasmic, 1% (23) OM, 0.5% (12) secreted, whereas 15% (285) proteins remained with undefined localization (**Figure 1A** and **Table S1**).

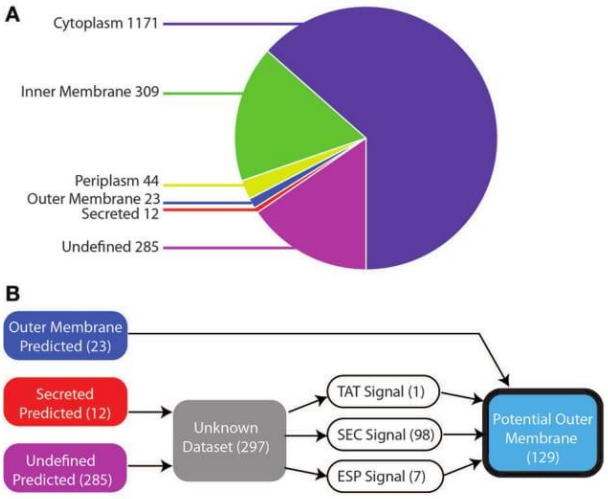

FIGURE 1 | Bioinformatic localization prediction. **(A)** Localization of *V. parvula* proteome predicted by PSORT, CELLO, SOSUI, LipoP, TMHMM, and BOMP (see text for details). **(B)** Strategy for refining prediction for the sequences with undefined localization (see text for details).

The 23 proteins robustly predicted to be in the OM include many known OM components such as TonB-dependent receptors, OmpM, components of the BAM/TAM machinery, TolC, and OmpA (**Table S1**). Although localization prediction methods were able to identify these OM components, they have limitations due to training on datasets of known OM proteins, most of which are from the phylogenetically distant Proteobacteria. This could prevent proper prediction of many proteins and fail to find novel OM proteins. Indeed, they failed to correctly identify the experimentally characterized trimeric autotransporters, which were predicted as secreted or had undefined localization (**Table S1**).

As such, we more thoroughly investigated the 297 proteins with undefined prediction or predicted as secreted, as they might contain additional OM proteins. As shown in **Figure 1B**, to do so, we constructed a bioinformatic filter to sort these 297 proteins. Briefly, we identified the proteins that contained a signal sequence for translocating through the IM, by checking for the three IM transport mechanisms: SEC, ESP (Extended signal peptide), and TAT with seven different programs (**Figure 1B**; see Section Materials and Methods). Hundred and ninety-one proteins lacked any recognizable signal sequence and were excluded. The remaining 106, which included all of the previously missed trimeric autotransporters, were combined with the 23 predicted OM proteins to provide a final dataset of 129 potential outer membrane proteins (**Table S2**).

Subcellular Fractionation

To validate our *in silico* prediction and to firmly identify OM proteins, we carried out a proteomic analysis of the OM of *V. parvula* DSM 2008 (see Section Materials and Methods for details). We performed three extractions: an outer membrane (OM) extract, a surface exposed (SE) extract, and a control whole cell (WC) protein extract. For all three extractions, *V. parvula* was grown anaerobically in BHIL medium in triplicate and the extraction was performed during the exponential growth phase. For OM protein extraction, we performed French press lysis to produce vesicles and isolated the OM vesicles with a chaotropic agent. To identify SE proteins we used a cell impermeable substrate, PEG-Biotin, which would label any primary amine exposed to the external environment of the cell. For each of the three extractions, only proteins that were present in all three biological replicates were considered for further analysis.

To identify proteins that were unique or shared by the different fractions we constructed a Venn diagram that is depicted in **Figure 2**. The WC control extract contained 1,342 proteins, while the OM fraction contained 990 proteins, and the SE fraction contained 849 proteins. Many false positives, including ribosomal proteins, were present in all three fractions (**Table S1**). This is a known product of the high sensitivity of MS analysis (Pocsfalvi et al., 2016).

Among the 1,342 proteins present in the WC extract, 696 were also detected in both the OM and SE fractions (**Figure 2**). Two hundred and sixty-two were detected in the WC and OM fractions but not in the SE. These proteins probably lack an exposed primary amine that would make them available to the biotinylation substrate. Similarly, 152 proteins were detected in the WC and SE fractions but not the OM fraction. These included many cytoplasmic or IM components and are therefore possible contaminants (**Table S1**). Finally, we found 32 proteins unique to the OM extract, and one protein unique to the SE extract. These discrepancies may be due to the enrichment of specific proteins by the OM and SE extraction protocols, combined with a too low concentration for detection in the WC control sample.

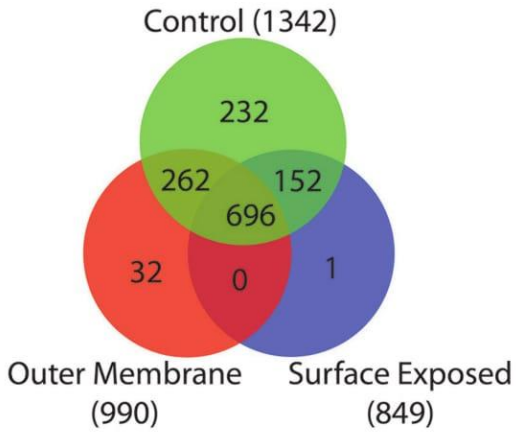

FIGURE 2 | Venn diagram of peptides detected from the three extraction methods. Overlap of peptides detected in the control sample (Green), OM extraction (Red), and SE extraction (Blue).

For the remainder of the analysis we only considered proteins as potential OM if they were detected in both the OM and the WC control samples, while we considered proteins as potential SE if they were present in all three samples, leading to 958 proteins (WC + OM = 262, WC + OM + SE = 696). In order to reduce false positives from this dataset, we crossed it with the 129 potential OM proteins obtained by our bioinformatics prediction (**Table S2**). This led to a final list of 78 OM proteins (**Table 1** and **Table S3**) that we annotated and created a graphical schematic representation in a *V. parvula* cell as shown in **Figure 3**. Identification of these OM proteins markedly extends previous inferences on the nature of the cell envelope of Negativicutes (Antunes et al., 2016) and provides information on the role of OM systems in the lifestyle of *Veillonella*.

Additionally, among the 1,844 predicted proteins of *V. parvula* roughly one quarter of them (502) were not expressed in our growth conditions or were not detected by MS due to their low concentration. As an example, we previously detected a conserved genomic locus coding for all components of a Type 4 pilus (T4P), including an OM secretin (Antunes et al., 2016). This cluster is similar to characterized Type 4A pili, which are known to be responsible for twitching motility (Pelicic, 2008) and natural competence (Knapp et al., 2017). However, we could only detect three peptides of PilM in one of the three WC control extraction replicates and three peptides of the component PilA in one of the three OM extraction replicates (**Table S4**). All remaining T4P components were absent in all of our samples, demonstrating the absence of their production in our growth conditions. Nonetheless, a T4P and the other undetected proteins may be produced in other conditions such as *in vivo* or during biofilm formation. Indeed, other species of *V. parvula* have shown differential competence capabilities in different media; this is likely related to T4P expression (Knapp et al., 2017).

An Ancestral β-Barrel Assembly Machinery, BAM/TAM, Is Present and Functional in the OM of *V. parvula*

The detection of OM proteins implies the existence of a functional system to insert them in the OM. The presence of a peculiar and probably ancestral BAM/TAM machinery has been inferred in *Negativicutes* and other bacteria (Webb et al., 2012; Heinz et al., 2015; Antunes et al., 2016). This system putatively contains three components, the BamA/TamA assembly barrel (Omp85), the TAM IM/periplasmic chaperone (TamB), and the BAM periplasmic chaperone (Skp). It has been hypothesized that this ancestral BAM/TAM machinery functions as both BAM and TAM in diderm Firmicutes (Heinz et al., 2015; Antunes et al., 2016). This hypothesis is supported by the detection of only a single Omp85 homolog in each genome and the fact all these genes form a conserved genomic cluster, suggesting functional linkage. Indeed, a recent study on TamB from *Borrelia burgdorferi* demonstrated that it interacts with BamA composing a BAM/TAM system (Iqbal et al., 2016).

Twenty-one of the 23 robustly predicted OM proteins were predicted as β-barrels, of which 17 were detected in our OM. This provides the first experimental evidence that the BAM/TAM

TABLE 1 | 78 OM proteins identified by bioinformatic localization and proteomic analysis. Order follows discussion in text.

Gene ID	Gene description	Robust prediction	WC	OM	SE
Vpar_0526	TamB	OM	+	+	+
Vpar_0527	BamA/TamA	OM	+	+	+
Vpar_0530*	Skp (OmpH)	Unclear	+	+	+
Vpar_1840	Skp (OmpH)	Unclear	+	+	+
Vpar_1106	Porin	Unclear	+	+	+
Vpar_0227	OmpM	OM	+	+	+
Vpar_0555	OmpM	OM	+	+	+
Vpar_0556	OmpM	OM	+	+	+
Vpar_0557	OmpM	OM	+	+	+
Vpar_0467	OmpA	OM	+	+	+
Vpar_0647	LptD	OM	+	+	+
Vpar_0548	LptA	Unclear	+	+	+
Vpar_0645	Uncharacterized protein	Unclear	+	+	+
Vpar_0646	Uncharacterized protein	Unclear	+	+	+
Vpar_0393	Uncharacterized protein	OM	+	+	+
Vpar_0051	Trimeric autotransporter adhesin	Secreted	+	+	+
Vpar_0052	Trimeric autotransporter adhesin	Secreted	+	+	+
Vpar_0464	Trimeric autotransporter adhesin	Unclear	+	+	+
Vpar_0048	Trimeric autotransporter adhesin	Unclear	+	+	+
Vpar_0042	Trimeric autotransporter adhesin	Unclear	+	+	+
Vpar_0046	Trimeric autotransporter adhesin	Unclear	+	+	+
Vpar_1413	TpsA	Unclear	+	+	+
Vpar_0041	Trimeric autotransporter adhesin	Unclear	+	+	+
Vpar_0100	Trimeric autotransporter adhesin	Unclear	+	+	+
Vpar_1664	Trimeric autotransporter adhesin	Unclear	+	+	+
Vpar_0053	Hemagglutinin domain protein	Unclear	+	+	+
Vpar_0330	Autotransporter	OM	+	+	+
Vpar_0298	Autotransporter-ShdA like	Secreted	+	+	
Vpar_1653	S-layer domain protein	Unclear	+	+	+
Vpar_1654	S-layer domain protein	Unclear	+	+	+
Vpar_1655	Uncharacterized protein	Unclear	+	+	+
Vpar_0074	TonB family protein	OM	+	+	+
Vpar_0061	TonB family protein	OM	+	+	+
Vpar_0065	TonB family protein	OM	+	+	+
Vpar_0066	TonB family protein	OM	+	+	+
Vpar_0719	TonB family protein	OM	+	+	+
Vpar_0525	TolC	OM	+	+	+
Vpar_1367	TolC	OM	+	+	+
Vpar_1003*	RND family efflux pump	Unclear	+	+	+
Vpar_0011*	RND family efflux pump	Unclear	+	+	
Vpar_1641*	Efflux transporter, RND family, MFP subunit	Unclear	+	+	
Vpar_0270	Peptidase: PepSY family	Secreted	+	+	+
Vpar_0057	Peptidase M48 family	Unclear	+	+	+
Vpar_0246	Peptidase: PepSY family	Unclear	+	+	+
Vpar_0412	Uncharacterized protein	Unclear	+	+	+
Vpar_0605	Peptidoglycan binding protein	Unclear	+	+	+
Vpar_1232*	Uncharacterized protein	Unclear	+	+	+
Vpar_1579	ATPase involved in DNA repair-like protein	Unclear	+	+	+
Vpar_1589*	Unknown Protein containing DUF3829 domain	Unclear	+	+	+
Vpar_0136	Unknown protein containing copper amine oxidase domain	Unclear	+	+	+
Vpar_0305	Uncharacterized protein	Unclear	+	+	+
Vpar_0469	Peptidase: PepSY family	Unclear	+	+	+
Vpar_0473	Unknown Protein containing DUF541 domain	Unclear	+	+	+
Vpar_0519	Unknown Protein containing DUF2233 domain	Unclear	+	+	+
Vpar_0521	Unknown protein	Unclear	+	+	+
Vpar_0562	Unknown protein containing DUF1421 domain	Unclear	+	+	+
Vpar_1148	Uncharacterized protein	Unclear	+	+	+
Vpar_1276	Uncharacterized protein	Unclear	+	+	+
Vpar_1760*	Unknown protein containing DUF3829 domain	Unclear	+	+	+
Vpar_1516	Uncharacterized protein	Unclear	+	+	+
Vpar_0809	Peptidase M23	Unclear	+	+	+
Vpar_1526	Uncharacterized protein	Unclear	+	+	+
Vpar_1767*	TPR repeat-containing protein	Secreted	+	+	
Vpar_0144	Hypothetical: MAEBL like protein	Unclear	+	+	
Vpar_0937*	VanW like protein	Unclear	+	+	
Vpar_0945*	Uncharacterized protein	Unclear	+	+	
Vpar_0965	Glycoside hydrolase family 18	Unclear	+	+	
Vpar_1765*	Unknown protein containing DUF3829 domain	Unclear	+	+	
Vpar_0828	HI0933 family protein	Unclear	+	+	
Vpar_0593	SpoIID/LytB domain protein	Secreted	+	+	+
Vpar_0260*	Periplasmic binding protein	Unclear	+	+	+
Vpar_0765*	Periplasmic binding protein	Unclear	+	+	+
Vpar_0410*	Periplasmic binding protein	Unclear	+	+	+
Vpar_1045*	Beta-lactamase domain protein	Unclear	+	+	+
Vpar_1188	Hydrogenase (NiFe) small subunit HydA	Unclear	+	+	+
Vpar_1773	Beta-N-acetylhexosaminidase (EC 3.2.1.52)	Unclear	+	+	+
Vpar_0358*	Periplasmic binding protein	Unclear	+	+	
Vpar_0764*	Periplasmic binding protein	Unclear	+	+	

Predicted lipoproteins that may be anchored to the inner membrane.

system is functional in *V. parvula*. Moreover, we detected the major OM component BamA/TamA (Vpar_0527), as well as TamB (Vpar_0526) in both our SE and OM fractions (**Figure 3**, **Table 1**, and **Table S3**). While the presence of the OM component BamA/TamA was expected, that of TamB is surprising, as TamB in other species is known to be periplasmic and attached to the IM. One can speculate that in *V. parvula* an entire complex containing BamA/TamA and TamB may be captured in the OM fraction. However, this fails to account for the strong OM prediction of TamB by localization software PSORT, CELLO, and SOSUI (**Table S1**), which indicates that some OM characteristics must be present within the *V. parvula* TamB; unfortunately

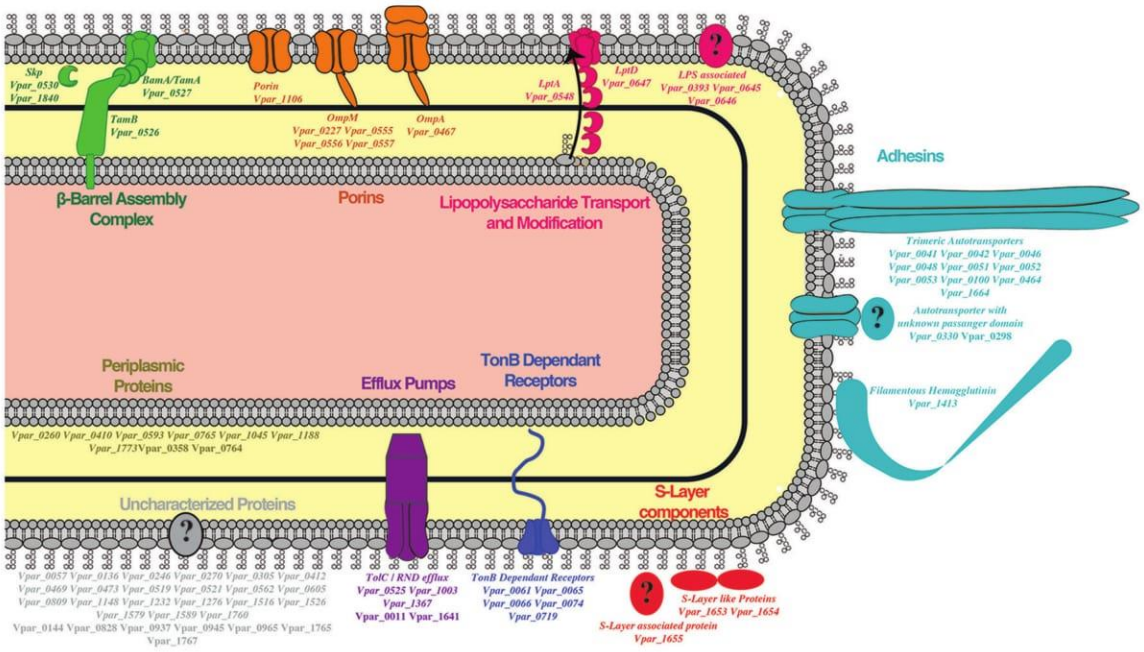

FIGURE 3 | Outer membrane proteome schematic demonstrating the different systems detected by our methods. Proteins found in the OM are in bold while SE detection is indicated by bold and italics.

we cannot determine what these characteristics are, due to the machine learning nature of the programs. Interestingly, TamB of *E. coli* (NP_418642.1) is also predicted to be in the OM by two out of the three programs (not shown), yet its IM/periplasmic localization has been firmly established through protease shaving (Selkrig et al., 2012). Discrepancies also exist in previous fractionation mass-spectrometry studies, as TamB was found in higher concentration in the OM than in the IM in *E. coli* (Martorana et al., 2014). Together, these data suggest that TamB may actually be both in the OM and the IM, or span the two membranes.

Four genes potentially encode the chaperone Skp in the *V. parvula* DSM 2008 genome (*vpar_0528-30, vpar_1840*), three of which are present in the diderm cluster immediately downstream of the BamA/TamA gene (Antunes et al., 2016). All four Skp proteins were detected in our OM fraction. However, two of them were excluded by our localization prediction procedure because of their strong prediction as periplasmic proteins (**Table S1**). Presence of Skp in the OM fraction is not unusual as Skp in *Salmonella* was originally thought to be in the OM, hence the old name OmpH (Thome and Müller, 1991). Furthermore, it has been shown experimentally that this protein interacts strongly with both the BAM system and OmpA, by functioning as a chaperone helping insertion of proteins in the OM (Selkrig et al., 2014).

Porins Are a Fundamental Component of *Veillonella*'s Outer Membrane

A key component of bacterial OM are porins, which are necessary for passive diffusion of small molecules (Galdiero et al., 2012). The genome of *V. parvula* contains genes encoding a typical porin (*vpar*_1106) consisting entirely of a β-barrel, four copies of OmpM (*vpar_0227, vpar_0555-7*) which contains a porin domain and a periplasmic SLH domain, and a homolog of the pathogenicity factor OmpA (*vpar_0467*) containing typical OmpA and beta-barrel domains (Confer and Ayalew, 2013).

We previously described the presence of three OmpM encoding genes in tandem within the diderm cluster of *V. parvula* (*vpar_0555-7*) (Antunes et al., 2016). During the course of this analysis we discovered a fourth (*vpar_0227*) at a different locus. The three OmpM proteins found within the diderm cluster (Vpar_0555-7) are among the most abundant proteins in the OM fraction by both quantification of our LC-MS/MS data (**Table S5**) and SDS-PAGE band excision MS (Bands 4 and 5) (**Table S6**). This correlates well with the function as an OM tether, as in *E. coli* Braun's lipoprotein is the most numerous protein in the cell (Braun, 1975). Furthermore, this supports the hypothesis that all Negativicutes utilize OmpM as a form of attachment (Kojima and Kamio, 2012). Moreover, this is consistent with a transcriptomics study of three *Veillonella* species from the mouth, which found all three copies of OmpM to be the most abundant OM transcripts (Do et al., 2015), further highlighting the importance of these proteins (Kojima et al., 2016).

We detected all four copies of OmpM in both the OM and the SE fractions (**Figure 3**, **Table 1**, and **Table S3**). Because of OmpM's abundance and dominant role in the OM, an exposed area of *Veillonella*'s OmpM may be involved in adhesion and biofilm formation. Previous bioinformatics studies have suggested that OmpM in the Negativicute *Selenomonas ruminantium* may be SE and interact with other bacteria (Kojima

et al., 2011). We previously wondered why there are multiple copies of the OmpM tether in diderm Firmicutes (Antunes et al., 2016); indeed, Braun's lipoprotein is usually in single copy, with the exception of *Salmonella's* from a recent gene duplication event (Sha et al., 2004). Because the four OmpM proteins are not identical, it is possible that they each provide a different surface for adhesion and/or immune evasion.

We also detected an OmpA homolog (Vpar_0467) in the OM and SE fractions. OmpA is an important protein in the OM of bacteria as it has strong pathogenic roles involving cellular invasion, adhesion, and host cell evasion (Confer and Ayalew, 2013) and therefore may be involved specifically in pathogenicity of *V. parvula*. Consistent with this hypothesis, a search for OmpA homologs in other Negativicutes revealed its presence in a few other human-related members of the family Veillonellaceae (**Figure 4**).

A Complete LPS Transport System

As we previously reported (Antunes et al., 2016), Negativicutes have the genetic potential to make and transport LPS to the OM. Most LPS biosynthesis genes are embedded in the diderm cluster in Negativicutes genomes, including that of *V. parvu* (**Figure 5A**). We were previously unable to identify the OM flippase components LptD or LptE, nor any component of potential O-antigen biosynthesis (Antunes et al., 2016). Here, an LptD homolog (Vpar_0647) was detected in the OM and SE fractions and identified as being an OM protein by localization prediction software (**Figure 3**, **Table 1**, and **Table S1**). This potential LptD encoding gene is located outside of the diderm cluster in a different region of the genome (**Figure 5B**). Interestingly, two proteins encoded by genes upstream of *lptD* (*vpar_0645* & *0646*) were both found in the OM and SE fractions (**Figure 5B** and **Table S1**). In addition, *vpar_0646* and *lptD* are

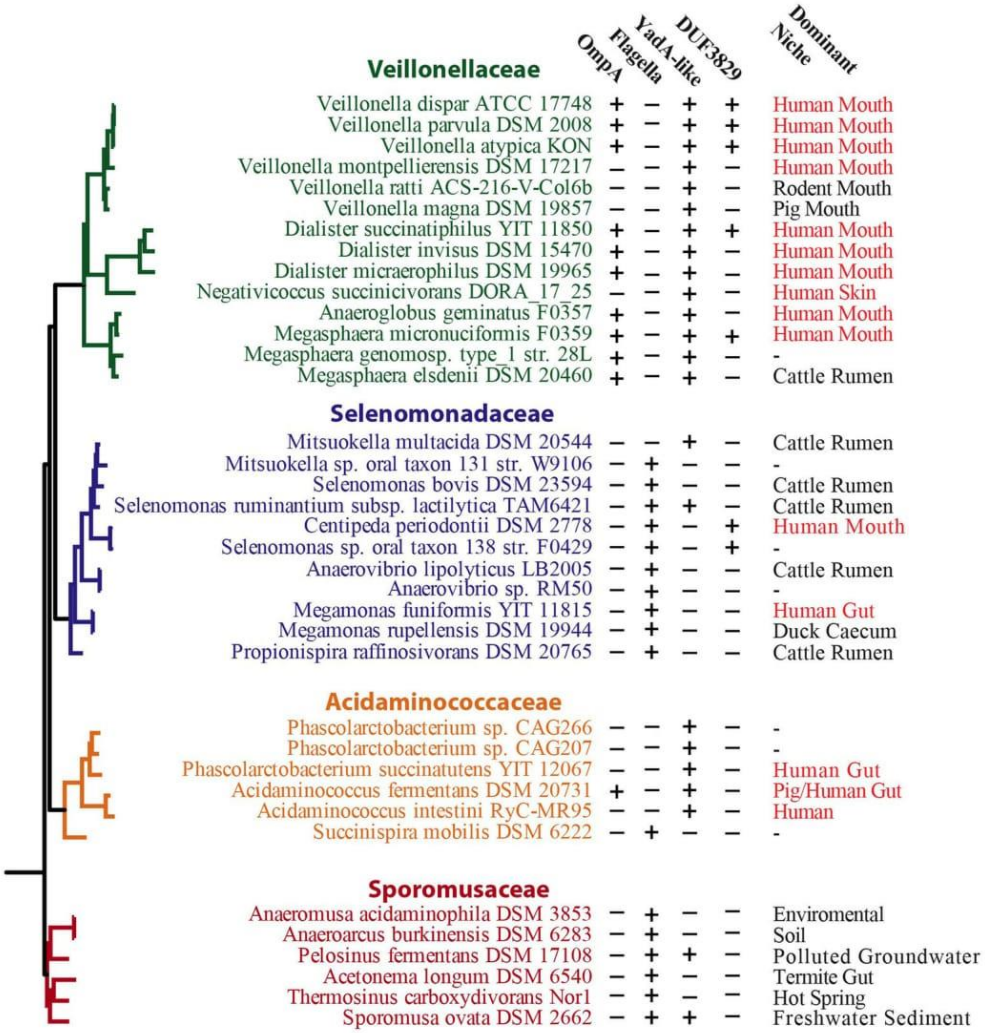

FIGURE 4 | Distribution of selected protein-coding genes in select proteins among Negativicutes. Schematic tree based on phylogeny of the Negativicutes from previous analysis (Antunes et al., 2016). Presence and absence of key genes are indicated by + and − symbols, respectively. The dominant niche of each strain taken from the available literature is also given. For discussion, see main text.

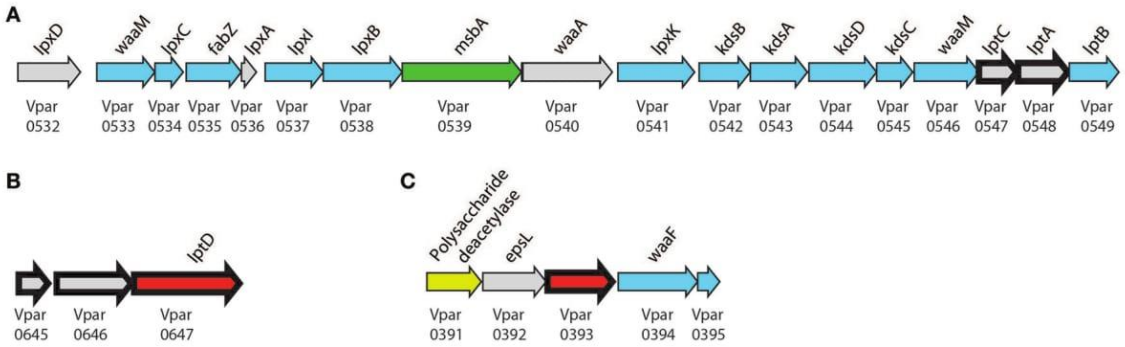

FIGURE 5 | Lipopolysaccharide biosynthesis clusters. Biosynthesis clusters found within *V. parvula* of **(A)** core biosynthetic and transport machinery, **(B)** conserved *lptD* cluster, **(C)** proposed LPS modification cluster. Bold arrows represent peptides detected in the OM. Localization is presented by color: Gray, unclear; Blue, cytoplasmic; Green, IM; Yellow, Periplasmic; and Red, OM. Genomic coordinates are **(A)**, CP001820.1:650916–670828 (+); **(B)** CP001820.1:780878–784136 (+); **(C)**, CP001820.1:477425–481382 (+).

conserved in sequence and synteny among all Negativicutes members (**Table S3**). We were unable to ascertain the function of Vpar_0645 & 0646 as we could find no conserved domains or clear homologs in sequence databases. It is possible that these proteins function with the Lpt complex of *V. parvula* or are functional equivalents of LptE, which caps the transport pore of LptD.

Finally, we detected the periplasmic components LptA (Vpar_0548) and LptC (Vpar_0547) in our OM and surface-exposed fractions. LptC was excluded from our final OM dataset because it did not pass the bioinformatics filter due to a lack of any discernible signal sequence. While we cannot exclude a contamination of periplasmic components in our OM fraction, it is possible that LptA and LptC were dragged to the OM fraction by strong interaction with LptD, thereby suggesting the existence of a complete Lpt complex in *V. parvula* similar to other studied bacteria (Villa et al., 2013).

O-Antigen in *V. parvula*?

Despite the presence of a complete pathway to make the lipid-A core of LPS, we previously could not predict if Negativicutes make O-antigen or not (Antunes et al., 2016), and the structure of *Veillonella*'s LPS is currently unknown. Here, we detected a protein in the OM fraction that may be used for further LPS processing (Vpar_0393). Although this protein has no conserved domains, we could tentatively infer its function based on the surrounding genomic region (**Figure 5C**). This region contains three genes probably involved in sugar anabolism (*vpar*_0391, 0392, 0394) and is well conserved only among the Negativicutes. The first protein, Vpar_0391, contains a CE4_SF domain; this domain is characteristic of polysaccharide deacetylases, such as PgaB or IcaB. The second protein, Vpar_0392, has the highest similarity to EpsL from a plasmid of *Lactococcus lactis,* which has no known function, yet is the final gene in the *eps* operon of exopolysaccharide biosynthesis (Forde and Fitzgerald, 2003). Although Vpar_0392 and EpsL are clear homologs, none of the surrounding genes bear any remarkable similarity. Vpar_0394 is annotated as WaaF, an enzyme that is known to be specific of diderm bacteria. It is responsible for O-antigen attachment to LPS by means of transferring a L-glycero-D-manno-heptose residue to the core oligosaccharide moiety of LPS. Mutation of this gene in *Salmonella* results in increase in sensitivity to antibiotics, detergents, and bile salts (Brooke and Valvano, 1996). Together, these data suggest that this could represent a new gene cluster responsible for O-antigen biosynthesis. To confirm this hypothesis and evaluate whether *V. parvula* LPS contains O-antigen we extracted and visualized its LPS on SDS PAGE (**Figure S1**). The extraction showed the characteristic ladder-banding pattern of O-antigen, however further characterization is necessary to determine *V. parvula* LPS structure and function.

V. parvula Has Many Potential Surface Exposed Adhesins

One of the most prevalent and important aspects of *Veillonella*'s lifestyle is its ability to form biofilms and interact with host cells or other members of the microbiota where it has been identified. Such interactions are mediated by adhesins and *Veillonella* has many of them (Hughes et al., 1992); these adhesins represent prolonged adaptation and evolution to the biofilm niche of *Veillonella*. By controlling the type and quantity of adhesins expressed it could prefer one substrate or cell and bind with high affinity. The presence of eight putative hemagglutinins belonging to the YadA-like family of trimeric autotransporters was recently reported in *V. atypica* OK5, and one of them (Hag1) was shown to be responsible for co-aggregation with resident flora in the oral cavity (Zhou et al., 2015).

Among the 12 trimeric autotransporters containing a typical YadA anchor domain in *V. parvula* DSM 2008 (**Figure S2**), we found 10 in all three samples (**Table 1**). These 10 trimeric autotransporters possess typical Yad_Head, Yad_Stalk, and Cter Beta-barrel Yad _anchor domains. They are likely adhesins of the hemagglutinin family (Bassler et al., 2015). Strikingly, six of these 10 trimeric autotransporters are located in a massive gene cluster (56 kb) (**Figures S2**, **S3A**). It is difficult to ascertain if this cluster is conserved in all Negativicutes and how many trimeric autotransporters it contains, as most of these genomes are in scaffolds. Trimeric autotransporters contain many repeated

regions, are modular in nature, contain many recent duplication events, and may get to be very large; as such, many Negativicutes scaffolds terminate in these genes thereby preventing synteny analysis. Two trimeric ATs are notable (**Figure S2A**): Vpar_0450 was not included in our final list of 78 OM proteins due to no detection in the OM fraction. It probably had too low of a concentration to be detected. Vpar_0045 is composed almost entirely of a Yad_anchor domain and thus is probably not functional and was not detected in any of our samples.

Interestingly, three trimeric autotransporter adhesins of *V. parvula* (Vpar_0041, Vpar_0100, Vpar_1664) possess a C-terminal SLH motif that is located downstream of the Beta-barrel YadA_anchor domain (**Figure S2**). This suggests a topology where the SLH domain is periplasmic and interacts with the peptidoglycan, similar to OmpM, to stabilize the trimeric configuration of these adhesins. We found this type of architecture between SLH and YadA anchor domains present in other autotransporters. However, it is restricted to the Negativicutes pointing to a unique feature potentially linked to their lifestyle (data not shown).

Upon further analysis, we noticed a distinct pattern of distribution for the presence of flagella and trimeric autotransporters in Negativicutes (**Figure 4**). The Veillonellaceae and Acidaminococcaceae contain YadA-like trimeric autotransporters and lack flagella, while the Selenomonadaceae and Sporomusaceae typically possess flagella and generally lack the YadA-like proteins, with only two exceptions (**Figure 4**). This peculiar distribution of flagella or adhesins between strains has not been observed previously and presents a unique case of evolution driven by lifestyle with the distinct presence of adhesion or motility. This observation provides a unique opportunity to study the transition between adhesion and motile life style, especially in *Mitsuokella multacida* DSM 20544, which represents an interesting specimen: it belongs to the Selenomonadaceae, which are typically flagellated, however it has lost its flagellum cluster and has acquired adhesins (**Figure 4**).

The *V. parvula* genome also contains two potentially functional filamentous hemagglutinin (Fha)-like proteins transported by the two-partner system (TPS) (**Figure S2B**): Vpar_0979/Vpar_0980 and Vpar_1413/Vpar_1414 as TpsA adhesin/TpsB transporter pair. We detected only the TpsA adhesin Vpar_1413 in the OM and SE fractions (**Figure 3**, **Table 1**, and **Table S3**). Its transporter Vpar_1414 was detected in the WC fraction but may have been present at too low concentration to be detected in the OM or SE fractions. The other system, Vpar_0979-Vpar_0980, was not detected in any fraction, including the whole protein fraction, suggesting an absence of its expression in our growing conditions. It may only be expressed upon contact with a host cell or when interacting with other microbial species. Furthermore, Vpar_0979 SEC secretion signal was only detected by one out of three programs and may represent a degenerate coding sequence.

In addition to trimeric and TPS systems that almost exclusively correspond to adhesins, *V. parvula* encodes classical autotransporters that contain a C-terminus β-barrel domain for insertion in the OM and an N-terminus passenger domain for function. Although some classical autotransporters are adhesins, within others the passenger domain may carry additional virulence functions such as protease or lipase domains (van Ulsen et al., 2014). We found four genes encoding such classical autotransporters within the genome (*vpar_0037, vpar_0298, vpar_0330, vpar_1322*) (**Figure S2C**). Vpar_0037 seems to be composed solely of a Beta-Barrel domain without any passenger domain or signal sequence; as such it is probably not functional. Vpar_1322 has a passenger domain with no detectable secondary or tertiary structural homology to known functional domains. Neither of these two autotransporters were detected in any of our samples suggesting they are not produced in our growth conditions. One that was found in the OM fraction, Vpar_0298, has a detectable homology to ShdA, the AidA-like adhesin of *Salmonella typhimurium*. Vpar_0330 is an autotransporter with no clear functional identification of its passenger domain, however we could find weak homology to a polysaccharide lyase-like protein using Phyre2 (see Section Materials and Methods); we detected Vpar_0330 both in our OM and SE fractions. Interestingly, *vpar_0330* was recently shown to be upregulated in *V. parvula* in a mouse tumor colonization model compared to *in vitro* growth, both with and without co-culture of *Pseudomonas aeruginosa* (Pustelny et al., 2015). Furthermore, it was highly upregulated in caries when compared to healthy teeth (Do et al., 2015). These data suggest that this autotransporter may be an important colonization factor in *Veillonella* and a priority target for further investigation.

Other Outer Membrane Features

Lipoproteins

Lipoproteins are found in most diderm bacteria and can be targeted to the outer membrane by the LOL transport machinery (Sutcliffe, 2010; Dowdell et al., 2017). We previously noted the absence of any LOL transport components (*lolA-E)* in any Negativicute genome (Antunes et al., 2016), however we find that 17 out of 78 OM proteins contain a lipoprotein signal sequence as predicted by LipoP and PRED-LIPO software (**Table 1** and **Table S3**). To resolve this incongruence, we further analyzed the MS data of OM proteins that contained a cysteine residue. We searched for any mono, di, or triacylglycerol with a saturated C16 or C18 fatty acid as a modification (Data not shown). No such modification was found on any of the proteins, supporting the hypothesis that *Veillonella* does not have OM-targeted lipoproteins. These data are not conclusive though, as our MS analysis was not designed specifically for this purpose, and further studies should be performed. If *V. parvula* is shown to possess OM lipoproteins, it may use an alternative pathway, which remains to be identified. Another possibility is that these proteins may be tethered to the IM with a lipid moiety and span the periplasmic space in addition to being in the OM.

Cell Ultrastructure

Although images of *Veillonella* are available in the literature, high quality electron microscopy has not been performed. We carried out ultrastructural analysis of *V. parvula* cells in two different ways: via chemical fixation (**Figures 6A,C,E**) and via high-pressure freezing of cells (**Figures 6B,D,F**). Subsequent

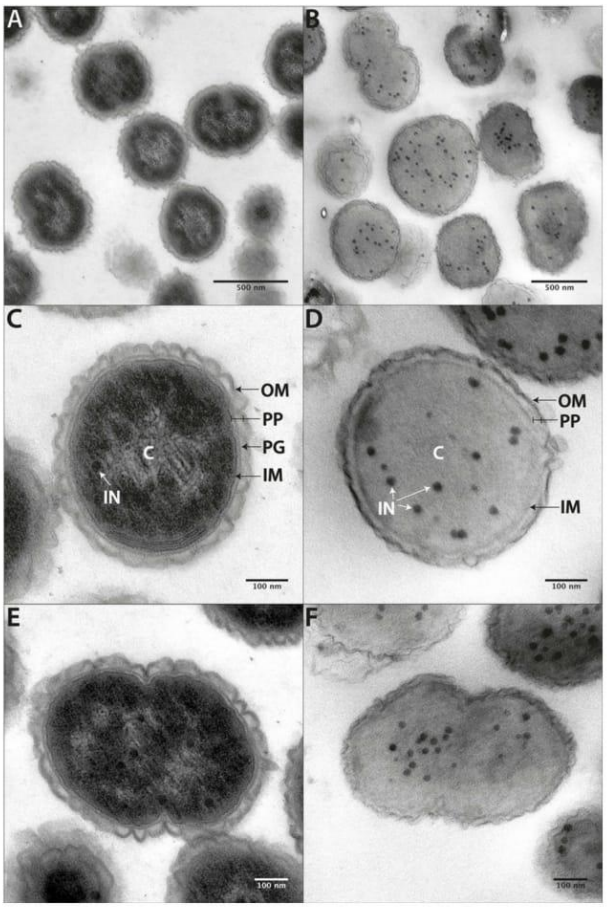

FIGURE 6 | Cell ultrastructure. Ultrathin sections of chemically fixed **(A,C,E)** or high-pressure frozen **(B,D,F)** *V. parvula* cells. At higher magnification the inner membrane (IM), the peptidoglycan (PG) containing periplasm (PP), and the slightly waved outer membrane (OM) become apparent. In chemically fixed cells, an electron brighter area within the periplasm most likely represents DNA. Electron dense circular structures (IN) can be seen within the homogenous cytoplasm of high-pressure frozen cells and, to a less extent also in chemically fixed cells.

resin embedding and ultrathin sectioning of *V. parvula* cells delivered a clear insight into the cell wall composition of this organism. Within the homogenous cytoplasm, electron dense particles are visible especially after high-pressure freezing (**Figures 6B,D,F**). In chemically fixed cells, an electron brighter area in the central part of the electron dense cytoplasm might show the DNA (genome). In both fixation variants, the cytoplasm is surrounded by the cytoplasmic membrane, which is again surrounded by a periplasm. Within this periplasm, peptidoglycan can sometimes be detected as electron dense line (**Figure 6C**). At the outermost part of the cell envelope, an outer membrane covers chemically fixed as well as high-pressure frozen cells. As the wavy appearance of the outer membrane is present after both fixation methods, this might either represent a special structural feature or a preparation artifact due to imperfect adjustment of the osmolarity of the post-fixation solutions to the original growth medium of the organism. The relative high thickness and in some cases fluffy appearance of the outer moiety of the outer membrane (**Figure S4**) is a strong indication for presence of LPS, although an S-layer cannot completely be ruled out (see below).

An S-Layer in *V. parvula*?

Surface layer proteins (S-layer proteins) are important cellular factors, in both monoderms and diderms. They are assembled as paracrystalline mono-layers on the cell surface of bacteria and archaea, are attachment sites for enzymes and/or substrates, or participate in interactions with abiotic surfaces or bacteria (Sara and Sleytr, 2000; Gerbino et al., 2015). We found two proteins containing three SLH contiguous domains (Vpar_1653 & Vpar_1654) in our OM and SE fractions (**Figure 3**, **Table 1**, and **Table S3**). Such domain architecture can be found in typical S-layer glycoproteins yet may also be associated to other functional domains such as hydrolases. Unfortunately, no other functional domains could be detected in Vpar_1653 and Vpar_1654. A classical S-layer has not been documented in *Veillonella*. A fluffy outer leaflet of the outer membrane was often visible, especially in high-pressure frozen cells (**Figure S4**). This may show the S-layer on top of the outer membrane, however it is more likely that this structure represents the LPS moiety. It is unlikely that a complete S-layer is produced, as these proteins are normally extremely abundant, but were 100 fold less abundant than OmpM in the OM fraction (**Table S5**).

Interestingly, *vpar_1653* and *vpar_1654* genes are in cluster along with a third gene (*vpar_1655*), which was also found in our OM and SE fractions (**Table 1**, **Table S3**). This third gene does not code for an S-layer protein and has no obvious function or conserved domains, other than one of unknown function (DUF4163). We could not find this three-gene cluster outside of *Veillonella* (**Table S3**), suggesting that it may have a genus-specific function. Further functional analysis is required to determine if *Veillonella* utilizes this gene cluster to make a typical S-layer or if it is used for enzymatic activity, interaction with surfaces and/or biofilm formation.

TonB-Dependent Transport

The TonB system is an OM transport mechanism present in most diderm bacteria, which has never been described in any Firmicute. Bacteria generally contain multiple TonB transport systems specifically associated with the acquisition of small molecules including vitamin B12 or metals such as iron (Krewulak and Vogel, 2011). We identified nine such TonB systems in the *V. parvula* genome, of which six were present in the OM and SE fractions (**Figure 3**, **Table 1**, and **Table S3**). Some of these systems must be functional to provide iron for the *V. parvula* haem cluster which has recently been characterized in *V. atypica* (Zhou et al., 2016). Interestingly, almost all of the operons encoding these TonB systems are located within a single genomic cluster downstream of the previously mentioned adhesion cluster (**Figure S3B**). Two other *tonB* operons are located at distant loci but were not found in our samples; specific molecules that were not present in our growing conditions might induce their expression.

Efflux Pumps

TolC is an outer membrane channel responsible for export of antibiotics and other toxic compounds from the cell. These systems are important in antibiotic resistance and in infection (Zgurskaya et al., 2011). Among the two TolC-ABC transporters and three similar RND type efflux pumps that are encoded in the genome of *V. parvula* DSM2008, we detected three in our OM and SE fractions (Vpar_0525, Vpar_1003, Vpar_1367) and two in only the OM fraction (Vpar_0011, Vpar_1641) (**Figure 3**, **Table 1**, and **Tables S1**, **S3**). Intriguingly, it was previously shown that the RND pump Vpar_1367 is upregulated in an infectious community responsible for caries (Do et al., 2015), pointing to the importance of such systems to scavenge essential molecules in *in vivo* conditions in a multispecies context.

Putative OM Proteins of Unknown Function

Among the list of 78 OM proteins, 28 proteins have no clear function (**Figure 3**, **Table 1**, and **Table S3**), many of which may correspond to previously undescribed systems and novel OM-related functions. We represented these unknown function proteins, as well as six unknown proteins previously described in other sections, with their detected functional domains in **Figure 7**.

Of special interest is Vpar_0521 (COG3064). We previously found it is well-conserved within the diderm Firmicute, and hypothesized its possible involvement in a yet unknown OM-related function (Antunes et al., 2016). Vpar_0521 contains two domains (**Figure 7**): an N-terminus SpoIVB and a C-terminal DUF1090. The N-terminus domain, SpoIVB, is an autoprotease (S5 peptidase) involved in sporulation in the

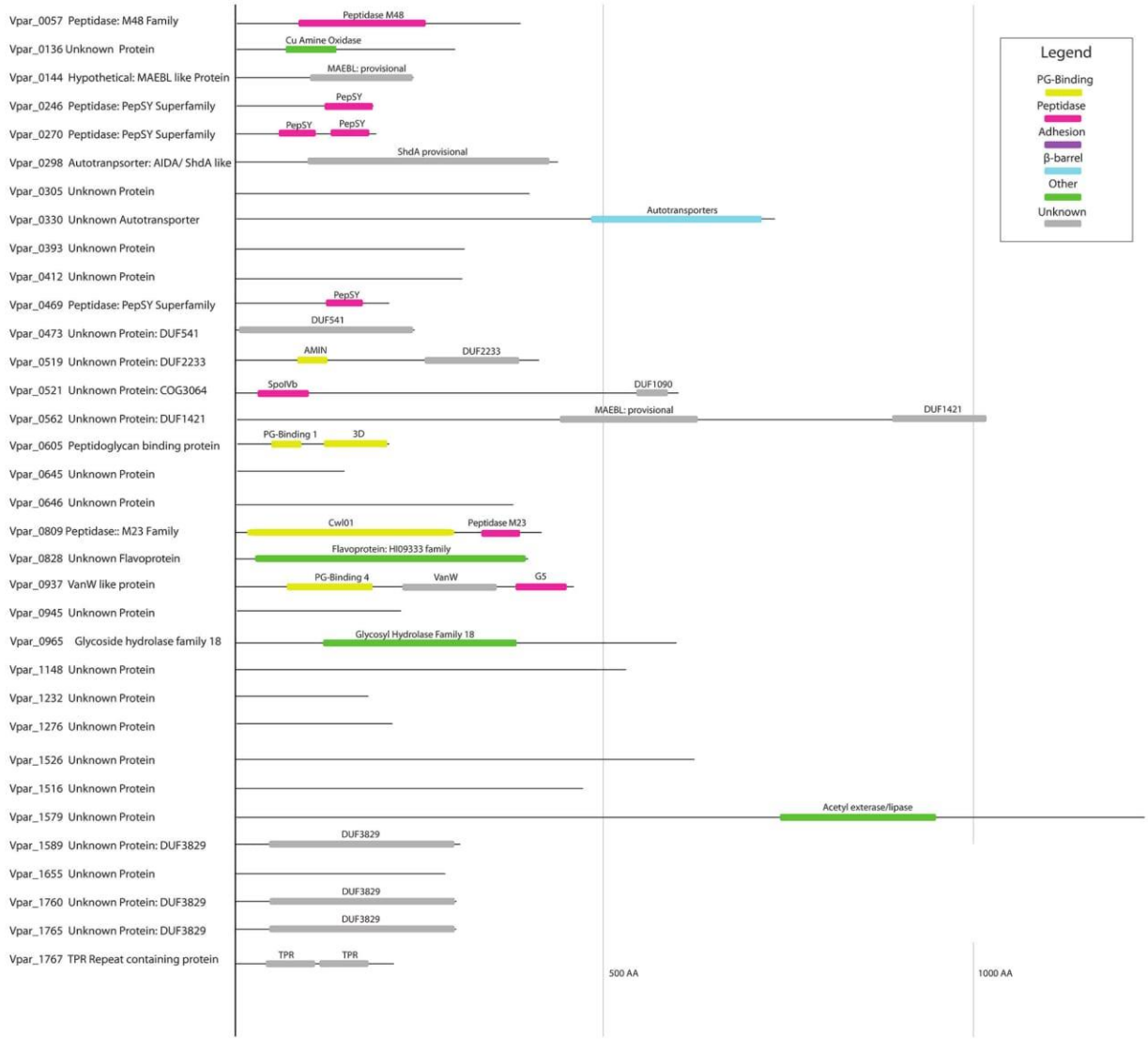

FIGURE 7 | Domain structure of unknown OM proteins. Protein domains identified in all poorly annotated OM proteins at the NCBI CDD database, and colored according to proposed function based on CDD annotation.

Firmicutes; it is normally synthesized in the forespore and transported to the interspace between the forespore membrane and the outer spore membrane, where it functions in signal transduction (Campo and Rudner, 2007). Because *V. parvula* is a non-sporulating Firmicute, the role of this SpoIVB domain remains to be determined. The other domain, DUF1090, is an uncharacterized domain found in all three domains of life, and predominantly present in Bacteria. For example, the *E. coli* YqjC protein only contains this domain. Only one study has been performed on the corresponding gene and it was found to be down regulated when the sigma factor *rpoS* gene was mutated (Zhang et al., 2012). In the Stepdb database that details the localization of all *E. coli* proteins (Orfanoudaki and Economou, 2014), YqjC is annotated as periplasmic. In some bacteria, this domain constitutes the C-terminal hydrophobic substrate-binding domain of the chaperone DnaK and it may have a similar function in Vpar_0521. These indications, and the conservation of Vpar_0521 in diderm Firmicutes, make this protein an important target of future study.

We found three proteins corresponding to the DUF3829 domain (*Vpar_1589, Vpar_1760, Vpar_1765*) (**Figure 7**). Although this domain is found in the distantly related Proteobacteria using a CDD search, the only close BLAST hits were from human-associated Negativicutes (**Figure 4**), more specifically the *Dialister*, *Centipeda*, *Selenomonas*, and *Veillonella*. We could not detect any homologs in *V. ratti* or any of the other rodents-associated *Veillonella* (Whitman, 2011). These unstudied proteins may be involved with interactions in the human niche, both for pathogenic and commensal lifestyles.

Three proteins contain one (Vpar_0246, Vpar_0469, Vpar_1597) or two (Vpar_0270) PepSY domains, and no other conserved domains. All proteins containing a PepSY domain were found in all three extractions except for Vpar_1597, which was not detected in any sample. This domain has not been extensively studied, yet it is known to possess a peptidase inhibitory function and is found throughout bacteria and some archaeal species (Yeats et al., 2004). The dominantly studied member of this family is YpeB, an inhibitor of the spore cortex lytic enzyme SleB in sporulating Firmicutes (Yeats et al., 2004). The four *V. parvula* OM PepSY proteins may function similarly to YpeB and inhibit peptidases of the M4 family or possess some unknown function.

CONCLUSIONS

In recent years we have learnt how important *Veillonella* is to the human microbiome, infection, and immune development (Whitman, 2011; Arrieta et al., 2015; Hirai et al., 2016). Moreover, these bacteria deserve to be studied not only for these characteristics, but also for the evolutionary questions they pose. This work represents the first proteomic characterization of a diderm Firmicute cell envelope and provides important information to guide further characterization. It confirms that the Negativicutes cell envelope has many aspects of the classical and well-studied Proteobacterial Gram-negative cell envelopes, such as LPS, TolC, OmpA, and other components, yet it also possesses many unique and potentially ancestral characteristics, such as the peculiar OM-PG attachment system, the BAM/TAM complex, as well as potentially new OM systems. The function of these systems needs to be further characterized through mutational and expression studies. Finally, our results are instrumental in increasing our understanding of the lifestyle of *Veillonella*, including its potential for biofilm formation and its role in infection and host-interaction.

AUTHOR CONTRIBUTIONS

DP prepared all samples and carried out bioinformatics analysis of genomic and proteomic data. MD, VH, and MM carried out mass spectrometry analysis. JF and AK carried out electron microscopy. CB supervised the study and analyzed the data. SG conceived and supervised the study, and analyzed the data. All authors participated in writing the manuscript.

FUNDING

This work was supported by grant ANR-16-CE12-0010-02 FIR-OM from "Agence Nationale de la Recherche," and PTR 39-16 from the Institut Pasteur "Programmes Transversaux de Recherche." DP is the recipient of a Ph.D. fellowship from the Institute Pasteur PPU program. CB is supported by the Institut Pasteur, the French Government Investissement d'Avenir Program Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant no. ANR-10-LABX-62-IBEID), Fondation pour la Recherche Médicale grant "Equipe FRM DEQ20140329508." SG acknowledges support by Investissement d'Avenir Grant "Ancestrome" (ANR-10-BINF-01-01).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01215/full#supplementary-material

Figure S1 | Glycostain of LPS. Thirteen and seventeen percentage SDS-PAGE loaded with identical samples and visualized with Pro-Q emerald glycostain. O-antigen producing positive control *E. coli* were loaded in lanes 1 and 3, while O-antigen negative extractions were loaded in lanes 2–4. Three biological replicates of *V. parvula* were loaded in lanes 5–7.

Figure S2 | Adhesins domain structure. Figure demonstrating the domain structure of all trimeric autotransporters **(A)**, two partner systems **(B)**, and autotransporters **(C)**.

Figure S3 | Trimeric autotransporter **(A)** and TonB **(B)** genomic clusters. Bold arrows represent peptides detected in the OM. Localization is presented by color: Gray, Unclear; Blue, Cytoplasmic; Green, IM; Yellow, Periplasmic; Purple, Secreted; and Red, OM.

Figure S4 | High-pressure frozen cell illustrating LPS. The ultrathin section of a high-pressure frozen cell shows the cell wall of *V. parvula* **(A)**. Beside the inner membrane (IM), the outer membrane (OM), and the peptidoglycan (PG) within the periplasm, the fluffy outer leaflet of the outer membrane is visible at higher magnification **(B)**. This might represent the LPS moiety. Bold arrows represent peptides detected in the OM.

Table S1 | Localization prediction and Mass spectrometry results for all proteins of *V. parvula* DSM 2008. Proteins predicted to be in the OM are in red.

Table S2 | Localization prediction and Mass spectrometry results for 129 potential OM proteins of *V. parvula* DSM 2008.

Table S3 | Localization prediction and Mass spectrometry results for 78 OM proteins of *V. parvula* DSM 2008.

Table S4 | Type 4 pili components *in silico* digestion and detection.

Table S5 | OM extraction gel excision Mass spectrometry results. Top 5 hits for each band are shown; Proteins predicted to be in the OM are highlighted.

Table S6 | IBAQ results for the OM extraction.

REFERENCES

Africa, C. W. J., Nel, J., and Stemmet, M. (2014). Anaerobes and bacterial vaginosis in pregnancy: virulence factors contributing to vaginal colonisation. *Int. J. Environ. Res. Public Health* 11, 6979–7000. doi: 10.3390/ijerph110706979

Antunes, L. C., Poppleton, D., Klingl, A., Dupuy, B., Brochier-Armanetd, C., Beloin, C., et al. (2016). Phylogenomic analysis supports the ancestral presence of LPS outer membranes in the Firmicutes. *Elife* 5:e14589. doi: 10.7554/eLife.14589

Arrieta, M.-C., Stiemsma, L. T., Dimitriu, P. A., Thorson, L., Russell, S., Yurist-Doutsch, S., et al. (2015). Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci. Transl. Med.* 7:307ra152. doi: 10.1126/scitranslmed.aab2271

Bagos, P. G., Nikolaou, E. P., Liakopoulos, T. D., and Tsirigos, K. D. (2010). Combined prediction of Tat and Sec signal peptides with hidden Markov models. *Bioinformatics* 26, 2811–2817. doi: 10.1093/bioinformatics/btq530

Bagos, P. G., Tsirigos, K. D., Liakopoulos, T. D., and Hamodrakas, S. J. (2008). Prediction of lipoprotein signal peptides in Gram-positive bacteria with a Hidden Markov Model. *J. Proteome Res.* 7, 5082–5093. doi: 10.1021/pr800162c

Bassler, J., Hernandez Alvarez, B., Hartmann, M. D., and Lupas, A. N. (2015). A domain dictionary of trimeric autotransporter adhesins. *Int. J. Med. Microbiol.* 305, 265–275. doi: 10.1016/j.ijmm.2014.12.010

Bendtsen, J., Nielsen, H., Widdick, D., Palmer, T., and Brunak, S. (2005). Prediction of twin-arginine signal peptides. *BMC Bioinformatics* 6:167. doi: 10.1186/1471-2105-6-167

Berven, F. S., Flikka, K., Jensen, H. B., and Eidhammer, I. (2004). BOMP: a program to predict integral beta-barrel outer membrane proteins encoded within genomes of Gram-negative bacteria. *Nucleic Acids Res.* 32, W394–W399. doi: 10.1093/nar/gkh351

Bladen, H. A., and Mergenhagen, S. E. (1964). Ultrastructure of Veillonella and morphological correlation of an outer membrane with particles associated with endotoxic activity. *J. Bacteriol.* 88, 1482–1492.

Braun, V. (1975). Covalent lipoprotein from the outer membrane of *Escherichia coli*. *Biochim. Biophys. Acta* 415, 335–377. doi: 10.1016/0304-4157(75)90013-1

Braun, V., and Rehn, K. (1969). Chemical characterization, spatial distribution and function of a lipoprotein (murein-lipoprotein) of the *E. coli* cell wall. The specific effect of trypsin on the membrane structure. *Eur. J. Biochem.* 10, 426–438. doi: 10.1111/j.1432-1033.1969.tb00707.x

Brooke, J. S., and Valvano, M. A. (1996). Biosynthesis of inner core lipopolysaccharide in enteric bacteria identification and characterization of a conserved phosphoheptose isomerase. *J. Biol. Chem.* 271, 3608–3614. doi: 10.1074/jbc.271.7.3608

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Campbell, C., Sutcliffe, I. C., and Gupta, R. S. (2014). Comparative proteome analysis of *Acidaminococcus intestini* supports a relationship between outer membrane biogenesis in Negativicutes and Proteobacteria. *Arch. Microbiol.* 196, 307–310. doi: 10.1007/s00203-014-0964-4

Campo, N., and Rudner, D. Z. (2007). SpoIVB and CtpB are both forespore signals in the activation of the sporulation transcription factor sigmaK in *Bacillus subtilis*. *J. Bacteriol.* 189, 6021–6027. doi: 10.1128/JB.00399-07

Chung, J. W., Ng-Thow-Hing, C., Budman, L. I., Gibbs, B. F., Nash, J. H. E., Jacques, M., et al. (2007). Outer membrane proteome of*Actinobacillus pleuropneumoniae*: LC-MS/MS analyses validate *in silico* predictions. *Proteomics* 7, 1854–1865. doi: 10.1002/pmic.200600979

Confer, A. W., and Ayalew, S. (2013). The OmpA family of proteins: roles in bacterial pathogenesis and immunity. *Vet. Microbiol.* 163, 207–222. doi: 10.1016/j.vetmic.2012.08.019

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* 13, 2513–2526. doi: 10.1074/mcp.M113.031591

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372. doi: 10.1038/nbt.1511

Davis, M. R., Goldberg, J. B., and Goldberg, J. B. (2012). Purification and visualization of lipopolysaccharide from Gram-negative bacteria by hot aqueous-phenol extraction. *J. Vis. Exp.* 2012:3916. doi: 10.3791/3916

Do, T., Sheehy, E. C., Mulli, T., Hughes, F., and Beighton, D. (2015). Transcriptomic analysis of three Veillonella spp. present in carious dentine and in the saliva of caries-free individuals. *Front. Cell. Infect. Microbiol.* 5:25. doi: 10.3389/fcimb.2015.00025

Dowdell, A. S., Murphy, M. D., Azodi, C., Swanson, S. K., Florens, L., Chen, S., et al. (2017). Comprehensive spatial analysis of the *Borrelia burgdorferi* lipoproteome reveals a compartmentalization bias toward the bacterial surface. *J. Bacteriol.* 199:e00658–16. doi: 10.1128/jb.00658-16

Erde, J., Loo, R. R. O., and Loo, J. A. (2014). Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.* 13, 1885–1895. doi: 10.1021/pr4010019

Euzeby, J. (1997). List of bacterial names with standing in nomenclature: a folder available on the internet. *Int. J. Syst. Bacteriol.* 590–592. doi: 10.1099/00207713-47-2-590

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., et al. (2017). InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* 45, D190–D199. doi: 10.1093/nar/gkw1107

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2015). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344

Forde, A., and Fitzgerald, G. F. (2003). Molecular organization of exopolysaccharide (EPS) encoding genes on the lactococcal bacteriophage adsorption blocking plasmid, pCI658. *Plasmid* 49, 130–142. doi: 10.1016/S0147-619X(02)00156-7

Galdiero, S., Falanga, A., Cantisani, M., Tarallo, R., Della Pepa, M. E., D'Oriano, V., et al. (2012). Microbe-host interactions: structure and role of Gram-negative bacterial porins. *Curr. Protein Pept. Sci.* 13, 843–854. doi: 10.2174/138920371121308001 2

Gerbino, E., Carasi, P., Mobili, P., Serradell, M. A., and Gómez-Zavaglia, A. (2015). Role of S-layer proteins in bacteria. *World J. Microbiol. Biotechnol.* 31, 1877–1887. doi: 10.1007/s11274-015-1952-9

Griffen, A. L., Beall, C. J., Campbell, J. H., Firestone, N. D., Kumar, P. S., Yang, Z. K., et al. (2012). Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *ISME J.* 6, 1176–1185. doi: 10.1038/ismej.2011.191

Hasegawa, K., Linnemann, R. W., Mansbach, J. M., Ajami, N. J., Espinola, J. A., Petrosino, J. F., et al. (2016). The fecal microbiota profile and bronchiolitis in infants. *Pediatrics* 51:474. doi: 10.1016/j.jemermed.2016.08.033

Heinz, E., Selkrig, J., Belousoff, M., and Lithgow, T. (2015). Evolution of the Translocation and Assembly Module (TAM). *Genome Biol. Evol.* 7:evv097. doi: 10.1093/gbe/evv097

Hirai, J., Yamagishi, Y., Kinjo, T., Hagihara, M., Sakanashi, D., Suematsu, H., et al. (2016). Osteomyelitis caused by Veillonella species: case report and review of the literature. *J. Infect. Chemother.* 22, 417–420. doi: 10.1016/j.jiac.2015.12.015

Hughes, C., Andersen, R., and Kolenbrander, P. (1992). Characterization of *Veillonella atypica* PK1910 adhesin-mediated coaggregation with oral Streptococcus spp. *Infect. Immun.* 60:1178.

Imai, K., Asakawa, N., Tsuji, T., Akazawa, F., Ino, A., Sonoyama, M., et al. (2008). SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in gram-negative bacteria. *Bioinformation* 2, 417–421. doi: 10.6026/97320630002417

Iqbal, H., Kenedy, M. R., Lybecker, M., and Akins, D. R. (2016). The TamB ortholog of *Borrelia burgdorferi* interacts with the β-barrel assembly machine (BAM) complex protein BamA. *Mol. Microbiol.* 102, 757–774. doi: 10.1111/mmi.13492

Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., and Krogh, A. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12, 1652–1662. doi: 10.1110/ps.0303703

Kall, L., Krogh, A., and Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server. *Nucleic Acids Res.* 35, W429–W432. doi: 10.1093/nar/gkm256

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053

Knapp, S., Brodal, C., Peterson, J., Qi, F., Kreth, J., and Merritt, J. (2017). Natural competence is common among clinical isolates of *Veillonella parvula* and is useful for genetic manipulation of this key member of the oral microbiome. *Front. Cell. Infect. Microbiol.* 7:139. doi: 10.3389/fcimb.2017.00139

Kojima, S., Hayashi, K., Tochigi, S., Kusano, T., Kaneko, J., and Kamio, Y. (2016). Peptidoglycan-associated outer membrane protein Mep45 of rumen anaerobe *Selenomonas ruminantium* forms a non-specific diffusion pore via its C-terminal transmembrane domain. *Biosci. Biotechnol. Biochem.* 80, 1954–1959. doi: 10.1080/09168451.2016.1194185

Kojima, S., and Kamio, Y. (2012). Molecular basis for the maintenance of envelope integrity in *Selenomonas ruminantium*: cadaverine biosynthesis and covalent modification into the peptidoglycan play a major role. *J. Nutr. Sci. Vitaminol. (Tokyo).* 58, 153–160. doi: 10.3177/jnsv.58.153

Kojima, S., Kaneko, J., Abe, N., Takatsuka, Y., and Kamio, Y. (2011). Cadaverine covalently linked to the peptidoglycan serves as the correct constituent for the anchoring mechanism between the outer membrane and peptidoglycan in *Selenomonas ruminantium*. *J. Bacteriol.* 193, 2347–2350. doi: 10.1128/JB.00106-11

Krewulak, K. D., and Vogel, H. J. (2011). TonB or not TonB: is that the question? *Biochem. Cell Biol.* 89, 87–97. doi: 10.1139/O10-141

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2014). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43, D222–D226. doi: 10.1093/nar/gku1221

Martorana, A. M., Motta, S., Di Silvestre, D., Falchi, F., Dehò, G., Mauri, P., et al. (2014). Dissecting *Escherichia coli* outer membrane biogenesis using differential proteomics. *PLoS ONE* 9:e100941. doi: 10.1371/journal.pone.0100941

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. (2015). PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* 44, D336–D342. doi: 10.1093/nar/gkv1194

Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41:e121. doi: 10.1093/nar/gkt263

Orfanoudaki, G., and Economou, A. (2014). Proteome-wide subcellular topologies of *E.* coli polypeptides database (STEPdb). *Mol. Cell. Proteomics* 13, 3674–3687. doi: 10.1074/mcp.O114.041137

Pelicic, V. (2008). Type IV pili: e pluribus unum? *Mol. Microbiol.* 68, 827–837. doi: 10.1111/j.1365-2958.2008.06197.x

Pocsfalvi, G., Stanly, C., Vilasi, A., Fiume, I., Capasso, G., Turiák, L., et al. (2016). Mass spectrometry of extracellular vesicles. *Mass Spectrom. Rev.* 35, 3–21. doi: 10.1002/mas.21457

Pustelny, C., Komor, U., Pawar, V., Lorenz, A., Bielecka, A., Moter, A., et al. (2015). Contribution of *Veillonella parvula* to *Pseudomonas aeruginosa*-mediated pathogenicity in a murine tumor model system. *Infect. Immun.* 83:417. doi: 10.1128/IAI.02234-14

Rosen, R., Amirault, J., Liu, H., Mitchell, P., Hu, L., Khatwa, U., et al. (2014). Changes in gastric and lung microflora with acid suppression. *JAMA Pediatr.* 168:932. doi: 10.1001/jamapediatrics.2014.696

Sara, M., and Sleytr, U. B. (2000). S-layer proteins. *J. Bacteriol.* 182, 859–868. doi: 10.1128/JB.182.4.859-868.2000

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342. doi: 10.1038/nature10098

Selkrig, J., Leyton, D. L., Webb, C. T., and Lithgow, T. (2014). Assembly of β-barrel proteins into bacterial outer membranes. *Biochim. Biophys. Acta* 1843, 1542–1550. doi: 10.1016/j.bbamcr.2013.10.009

Selkrig, J., Mosbahi, K., Webb, C. T., Belousoff, M. J., Perry, A. J., Wells, T. J., et al. (2012). Discovery of an archetypal protein transport system in bacterial outer membranes. *Nat. Struct. Mol. Biol.* 19, 506–510. doi: 10.1038/nsmb.2261

Sha, J., Fadl, A. A., Klimpel, G. R., Niesel, D. W., Popov, V. L., and Chopra, A. K. (2004). The two murein lipoproteins of Salmonella enterica serovar Typhimurium contribute to the virulence of the organism. *Infect. Immun.* 72, 3987–4003. doi: 10.1128/IAI.72.7.3987-4003.2004

Strati, F., Cavalieri, D., Albanese, D., De Felice, C., Donati, C., Hayek, J., et al. (2017). New evidences on the altered gut microbiota in autism spectrum disorders. *Microbiome* 5:24. doi: 10.1186/s40168-017-0242-1

Sutcliffe, I. C. (2010). A phylum level perspective on bacterial cell envelope architecture. *Trends Microbiol.* 18, 464–470. doi: 10.1016/j.tim.2010.06.005

Szijarto, V., Lukasiewicz, J., Gozdziewicz, T. K., Magyarics, Z., Nagy, E., and Nagy, G. (2014). Diagnostic potential of monoclonal antibodies specific to the unique O-antigen of multidrug-resistant epidemic *Escherichia coli* clone ST131-O25b:H4. *Clin. Vaccine Immunol.* 21, 930–939. doi: 10.1128/CVI.00685-13

Thein, M., Sauer, G., Paramasivam, N., Grin, I., and Linke, D. (2010). Efficient subfractionation of gram-negative bacteria for proteomics studies. *J. Proteome Res.* 9, 6135–6147. doi: 10.1021/pr1002438

Thome, B. M., and Müller, M. (1991). Skp is a periplasmic *Escherichia coli* protein requiring SecA and SecY for export. *Mol. Microbiol.* 5, 2815–2821. doi: 10.1111/j.1365-2958.1991.tb01990.x

Tocheva, E. I., Matson, E. G., Morris, D. M., Moussavi, F., Leadbetter, J. R., and Jensen, G. J. (2011). Peptidoglycan remodeling and conversion of an inner membrane into an outer membrane during sporulation. *Cell* 146, 799–812. doi: 10.1016/j.cell.2011.07.029

Tusher, V. G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.* 98, 5116–5121. doi: 10.1073/pnas.091062498

Tyanova, S., Temu, T., and Cox, J. (2016a). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* 11, 2301–2319. doi: 10.1038/nprot.2016.136

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., et al. (2016b). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 13, 731–740. doi: 10.1038/nmeth.3901

van Ulsen, P., Rahman, S., Jong, W. S. P., Daleke-Schermerhorn, M. H., and Luirink, J. (2014). Type V secretion: from biogenesis to biotechnology. *Biochim. Biophys. Acta* 1843, 1592–1611. doi: 10.1016/j.bbamcr.2013.11.006

Veillon, A., and Zuber, A. (1898). Sur Quelques microbes strictment anaerobies et leur role en pathologie. *Arch. Med. Exp.* 517–545.

Villa, R., Martorana, A. M., Okuda, S., Gourlay, L. J., Nardini, M., Sperandeo, P., et al. (2013). The *Escherichia coli* Lpt transenvelope protein complex for lipopolysaccharide export is assembled via conserved structurally homologous domains. *J. Bacteriol.* 195, 1100–1108. doi: 10.1128/JB.02057-12

Vizcaíno, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., et al. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, D447–D456. doi: 10.1093/nar/gkw880

Voss, B., Gaddy, J., McDonald, W., and Cover, T. (2014). Analysis of surface-exposed outer membrane proteins in *Helicobacter pylori*. *J. Bacteriol.* 196, 2455–2471. doi: 10.1128/JB.01768-14

Webb, C. T., Heinz, E., and Lithgow, T. (2012). Evolution of the ß-barrel assembly machinery. *Trends Microbiol.* 20, 612–620. doi: 10.1016/j.tim.2012.08.006

Whitman, W. B. (2011). "Veillonellaceae" in *Bergey's Manual of Systematic Bacteriology, Vol. 3, The Firmicutes, 2nd Edn.*, eds. D. Vos, G. Garrity, D. Jones, N. Krieg, W. Ludwig, F. A. Rainey, K.-H. Schleifer, and W. Whitman (New York, NY: Springer), 1059–1119.

Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., et al. (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* 379, 466–469. doi: 10.1038/379466a0

Wilson, M. M., Anderson, D. E., Bernstein, H. D., McInnes, P., and Wang, L. (2015). Analysis of the outer membrane proteome and secretome of *Bacteroides fragilis* reveals a multiplicity of secretion mechanisms. *PLoS ONE* 10:e0117732. doi: 10.1371/journal.pone.0117732

Yeats, C., Rawlings, N. D., and Bateman, A. (2004). The PepSY domain: a regulator of peptidase activity in the microbial environment? *Trends Biochem. Sci.* 29, 169–172. doi: 10.1016/j.tibs.2004.02.004

Yu, C. S., Lin, C. J., and Hwang, J. K. (2004). Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* 13, 1402–1406. doi: 10.1110/ps.03479604

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608–1615. doi: 10.1093/bioinformatics/btq249

Zeikus, J. G., Hegge, P. W., Thompson, T. E., Phelps, T. J., and Langworthy, T. A. (1983). Isolation and description of *Haloanaerobium praevalens* gen. nov. and sp. nov., an obligately anaerobic halophile common to Great Salt Lake sediments. *Curr. Microbiol.* 9, 225–233. doi: 10.1007/BF01567586

Zgurskaya, H. I., Krishnamoorthy, G., Ntreh, A., and Lu, S. (2011). Mechanism and function of the outer membrane channel TolC in multidrug resistance and physiology of enterobacteria. *Front. Microbiol.* 2:189. doi: 10.3389/fmicb.2011.00189

Zhang, Y., Shi, C., Yu, J., Ren, J., and Sun, D. (2012). RpoS regulates a novel type of plasmid DNA transfer in *Escherichia coli*. *PLoS ONE* 7:e33514. doi: 10.1371/journal.pone.0033514

Zhou, P., Liu, J., Merritt, J., and Qi, F. (2015). A YadA-like autotransporter, Hag1 in *Veillonella atypica* is a multivalent hemagglutinin involved in adherence to oral streptococci, *Porphyromonas gingivalis*, and human oral buccal cells. *Mol. Oral Microbiol.* 30, 269–279. doi: 10.1111/omi.12091

Zhou, P., Li, X., and Qi, F. (2016). Identification and characterization of a haem biosynthesis locus in Veillonella. *Microbiology* 162, 1735–1743. doi: 10.1099/mic.0.000366

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

2.2 Outer Membrane Attachment in Bacteria

Within this analysis, there is one protein family from the diderm Firmicutes that continuously peaked my attention, OmpM, the OM tether. The interesting features of this protein are that it is so drastically different from the textbook example Lpp or Braun's lipoprotein. Lpp is a very small lipoprotein that is covalently attached to the PG while OmpM is a large porin that noncovalently interacts with putrescine modified peptidoglycan. I originally had thought that I would only need to map the distribution of OmpM as Lpp is a well-described protein and surely some laboratory must have done this analysis, but no this textbook example has only been studied in a limited set of γ-Proteobacteria.

When I started this analysis I was surprised to learn that the reason Lpp is only studied in the γ-Proteobacteria is due to it being a new adaptation of these bacteria. Conversely, OmpM was found to be widely distributed among the Terrabacteria, aka the Cyanobacteria, Atribacteria, Deinococcus-Thermus, Thermotoga, Dictyoglomi, and Armatimonadetes. This impressive distribution was surprising as it accounted for roughly half the bacterial clades, with a well-known monophyletic group. This did leave the question as to what the remainder of the bacteria used to attach their OM.

I conducted an intensive literature search and came up with two logical candidates, OmpA and the TolB/Pal complex. Both OmpA and Pal bind the meso-Dap residue of the PG with the same domain and both were suggested to have an OM attachment function. The difference lay in how they integrate into the OM; Pal is a lipoprotein while OmpM is a porin. After intensive queries for both systems, I found that OmpA had sporadic distribution and could not be a primary conserved mechanism, conversely, TolB/Pal was found in most clades that lacked OmpM, barring a few exceptions. This provided a mutual exclusion of systems with the Terrabacteria using OmpM and most other clades using TolB/Pal with no overlap.

As with any good scientific result, more questions arose than were solved. Why is this distribution split? I tried to answer this by mapping the peptidoglycan onto the existing tree and found that PG structure is consistently Meso-DAP among all bacteria that use a TolB/Pal system, while those that use OmpM have far more variability. I present here a draft of an article that is ready to publish pending minor analysis.

Hook, Line, and Sinker: an Ancient Division of Outer Membrane Tethering Systems in Bacteria

Daniel Poppleton[1], Pahagotis Adam[1] Christophe Beloin[2] and Simonetta Gribaldo[1].
[1]Unité de Biologie Moléculaire du Gène chez les Extrêmophiles, Département de Microbiologie, Institut Pasteur, Paris, France
[2]Unité de Génétique des Biofilms, Département de Microbiologie, Institut Pasteur, Paris, France

ABSTRACT
The mechanism of attachment of the outer membrane (OM) to the peptidoglycan (PG) is of fundamental importance in diderm bacteria, not only to maintain cell integrity but also for resistance to antibiotics and the production of OM vesicles. While the textbook example of OM attachment is *Escherichia coli*'s Lpp (Braun's lipoprotein), this mechanism is not universally distributed and therefore not representative of the diversity of diderm bacterial lineages. A second mechanism involving the OmpM protein has been described in diderm Firmicutes, Deinococci, Cyanobacteria, and Thermotogae. Moreover, two other mechanisms for OM attachment have been hypothesized (OmpA and a two-component TolB/Pal system).
The existence of multiple OM attachment mechanisms and the biological relevance of this mechanism poses the question of their distribution in Bacteria and their evolutionary history. To this aim, we have carried out a search for homologues of Lpp, OmpM, OmpA, and TolB/Pal in all major bacterial phyla. Lpp is patchily restricted to γ-Proteobacteria, suggesting a recent origin in this clade. OmpA is more widely distributed, but is also patchy, excluding the hypothesis that it represents the main alternative mechanism of OM-attachment. Surprisingly, the TolB/Pal and OmpM systems are much more widely present and follow a mutually exclusive distribution in Bacteria. In fact, the TolB/Pal system is largely present in the Gracilicutes. In contrast, OmpM is widely distributed among deep branching diderm lineages of the Terrabacteria. Finally, Fusobacteria and Spirochaetes might represent an intermediary step in the transition between these two systems, involving OmpA. Our results highlight an ancient division in OM attachment in Bacteria. OmpM might represent the ancestral mechanism, which was replaced by TolB/Pal in the ancestor of the Gracilicutes. This may be linked to PG modification, whereas Lpp was a more recent addition to the arsenal of OM tethering mechanisms in the γ-Proteobacteria. The reasons for these transitions may lie in the emergence of new PG modifications and environmental adaptations with changes in lifestyle. Beyond these important evolutionary implications, and the possibility of using OM-tethering systems as a marker for bacterial phylogeny, our results pave the way to further study of the vast diversity of uncharacterized bacterial cell envelopes.

INTRODUCTION
Early studies on *E. coli's* lipoproteins found the most abundant one, Lpp, at roughly 1 million copies per cell, to be responsible for tethering the OM to the peptidoglycan (PG) (Braun and Rehn, 1969). Lpp is attached to the OM via a lipid moiety, which integrates into the lipid bilayer (Figure 1) (Guo *et al.*, 2014). The small, 58 amino acids, molecule forms a trimer *in vivo* with one chain covalently bound to the diaminopimelate residue in the peptide cross bridge of PG (Shu *et al.*, 2000). This covalent bond is formed by the L, D transpeptidases, three of which have been identified in *E. coli*: ErfK, YcfS, and YbiS (Magnet *et al.*, 2008). *ybiS* if of special interest as it is directly upstream of *lpp* in many species. These peptidases all contain the YkuD active domain. Experimental evidence has been provided for homologs in *Pseudomonas aeruginosa*, *Salmonella enterocolitica*, and other γ-Proteobacteria (Ching and Inouye, 1985). Mutants of Lpp in *E. coli* have the noticeable phenotypic characteristics of a leaky and poorly attached OM. These characteristics are hypersensitivity to bile salts, some antibiotics and detergents, increased release of outer membrane vesicles (OMVs), leakage of periplasmic proteins and loss of OM asymmetry (Chen *et al.*, 2014). While the mechanism of OM attachment by Lpp is the paradigm written in introductory textbooks, the fact that it not essential in *E. coli* and even absent from many Proteobacteria, such as Neisseria (Adu-Bobie *et al.*, 2004) indicates that it is not the main system. In fact, in *E. coli* two other proteins have been suggested to have a potential role in OM attachment: OmpA and Pal (Schwechheimer *et al.*, 2014) (Figure 1). Pal is an OM lipoprotein that has been studied extensively as part of the 5-component Tol-Pal complex. The Tol-Pal complex has a wide variety of proposed functions. The Tol-Pal system is the target of certain colicins and bacteriocins that disrupt intermolecular interactions of this complex, thereby causing cell death (Kim *et al.*, 2014). One well-characterized feature of this complex is that disruption of any component results in increased OMV production and OM instability (Llamas *et al.*, 2000). Pal is lipidated in the same manner as Lpp and strongly associates, albeit noncovalently, with the PG and the periplasmic component TolB. Pal can also form an independent 2-component TolB/Pal system (Figure 1) (for a recent review see (Godlewska *et al.*, 2009)), and has an important role in OM stability (Cascales *et al.*, 2002). Indeed, overexpression of *pal* suppresses the phenotypes associated with *lpp* mutations (Cascales *et al.*, 2002). The PG-binding domain of Pal is named OmpA (henceforth referred to as D-OmpA to differentiate it from the protein of the same name) (Figure 1). The D-OmpA domain is found in many different proteins including Pal, the flagella stator protein MotB, the type VI secretion system protein DotU, and of course OmpA (Confer and Ayalew, 2013).

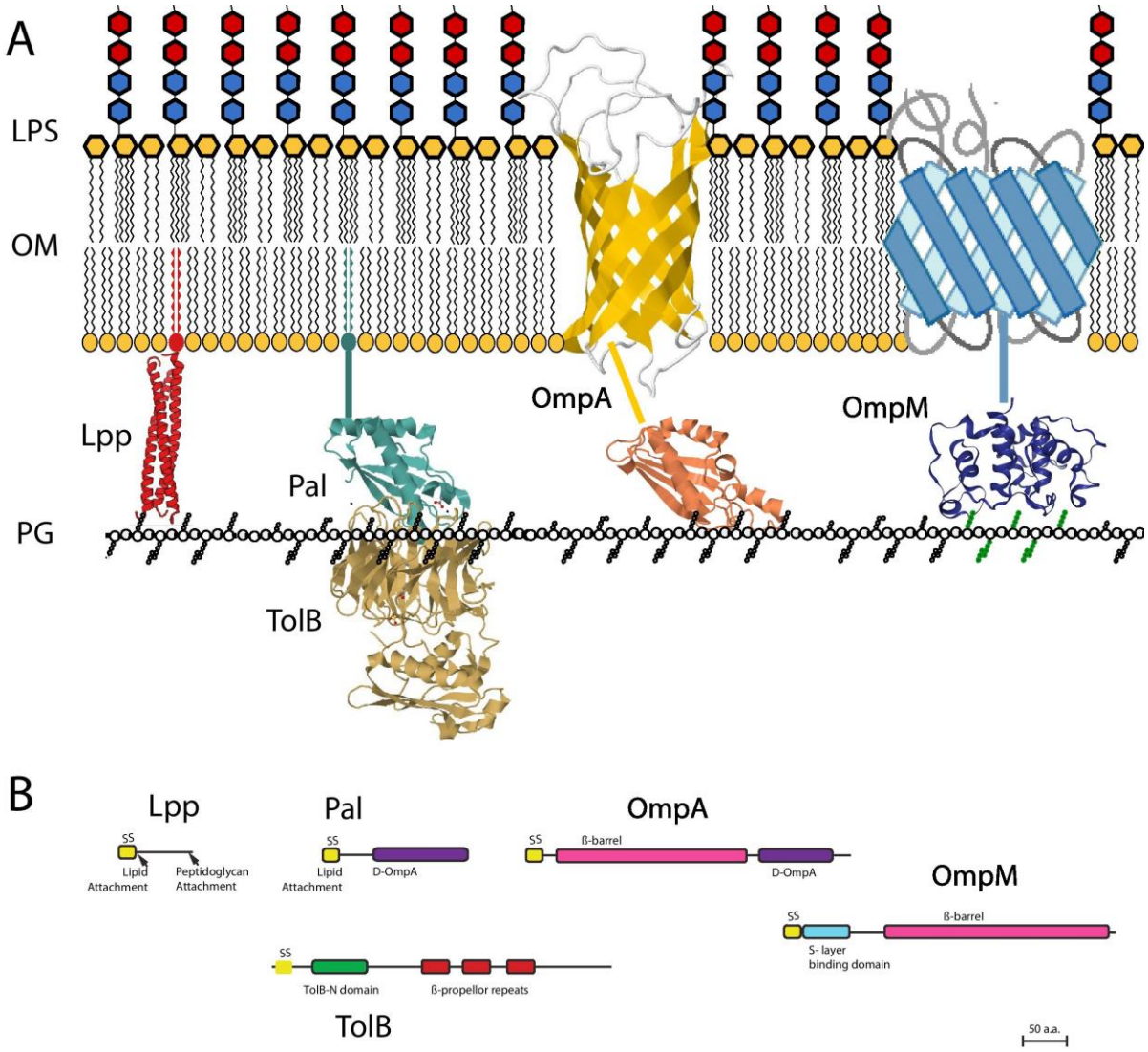

Figure 1: Schematic of the four different systems present in Bacteria for OM-attachment:
Lpp, TolB/Pal, OmpA, and OmpM (A), and the corresponding sequence characteristics (B) (see text for details). Structures are taken from the PDB database (Berman et al., 2000): Lpp (PDB ID: 1EQ7); Pal (PDB ID: 2HQS); OmpA (PDB ID: 2GE4); D-OmpA domain (PDB ID: 4ERH). OmpM (PDB ID: 3PYW).

OmpA is a pathogenicity factor that has many roles in immune system targeting and evasion, cell host adhesion and invasion, and as a porin (for a recent review, see (Confer and Ayalew, 2013). Like Lpp, OmpA is present in high abundance with around 100,000 copies per cell, it functions as a porin, and is surface exposed (Confer and Ayalew, 2013). Other than these roles, OmpA may also function in OM tethering. In fact, it can integrate into the OM with an N-terminal β-barrel domain, while it non-covalently interacts with the PG with the D-OmpA domain, similarly to Pal (Figure 1). Interestingly, it has been shown that OmpA can have an alternative conformation where it loses the PG attachment and functions solely as a porin (Sugawara *et al.*, 2006). Why this conformational change happens has not been determined, but it could allow quick solute exchange via a porin, or the loss of tethering could function as a mechanism of inducing OMVs. The function of OmpA in OM tethering has however never been fully demonstrated and has only been sparingly attributed, as this function has been predominantly studied in *E. coli* where Lpp is the dominant tether. In fact, while OmpA binds PG (Park *et al.*, 2012), its mutation does not always produce heightened OMV production like Lpp or Tol-Pal mutants (Wessel *et al.*, 2013) (Petrus *et al.*, 2012; Park *et al.*, 2012). Nonetheless, an *ompA* mutant was shown to have all of the characteristics of an OM-tether mutant, most notably membrane leakage, asymmetry, and sensitivity to SDS, cholate, and high osmolarity (Park *et al.*, 2012) as well as a 26 fold increase in OMV production (Schwechheimer *et al.*, 2014).

Outside of the Proteobacteria, where the Lpp, Pal, and OmpA systems have been extensively studied, there exists another form of OM attachment, OmpM. This mechanism is present in the Negativicutes, a lineage belonging to the Firmicutes that possess an OM with lipopolysaccharide (hereafter called “diderm Firmicutes”) (Whitman, 2011; Antunes *et al.*, 2016; Poppleton *et al.*, 2017). OmpM has been characterized in detail in the Negativicute *Selenomonas ruminantium* (for a recent review see (Kojima and Kamio, 2012)). It is functionally similar to OmpA in that it integrates into the OM via a β-

barrel and attaches to the PG via an SLH (S-like homology) domain at its N-terminus. The difference is in how this protein interacts with the PG; *Selenomonas* PG is modified with polyamines at the cross bridge where OmpM binds with the SLH domain (Figure 1). Although no mutational studies have been done due to a lack of genetic system in *Selenomonas*, polyamine depletion leads to a leaky OM, similar to Lpp and Pal mutants (Kojima *et al.*, 2011). We have recently suggested using *in silico* analysis and experimental characterization that OmpM is indeed the dominant form of attachment in all diderm Firmicutes, although some also possess a homologue of OmpA (Antunes et al. 2016; Poppleton *et al.*, 2017). Aside from diderm Firmicutes, OmpM has been identified as the main component of the OM of Deinococci (Misra *et al.*, 2015) and Cyanobacteria (Hansel *et al.*, 1998). The Thermotogae possess a two-component system that may be derived from OmpM: a β-barrel peptide (Ompβ) that constitutes the OM integration component and a second protein (Ompα) containing an SLH domain for PG binding (Lupas *et al.*, 1995).
The existence of different ways of OM attachment, and the biological relevance of this mechanism, poses the question of their actual distribution in Bacteria as well as their evolutionary history. The distribution of OmpA, OmpM, and Lpp has not been done. The distribution of the 5-component Tol/Pal system was looked at previously (Sturgis, 2001), however they only queried 22 genomes for the entire system and may have missed many important phyla. Therefore, we undertook a search of the four OM attachment systems in nearly 200 bacterial genomes representatives of major bacterial phyla, as well as a more detailed survey of Lpp homologues in the γ-Proteobacteria.

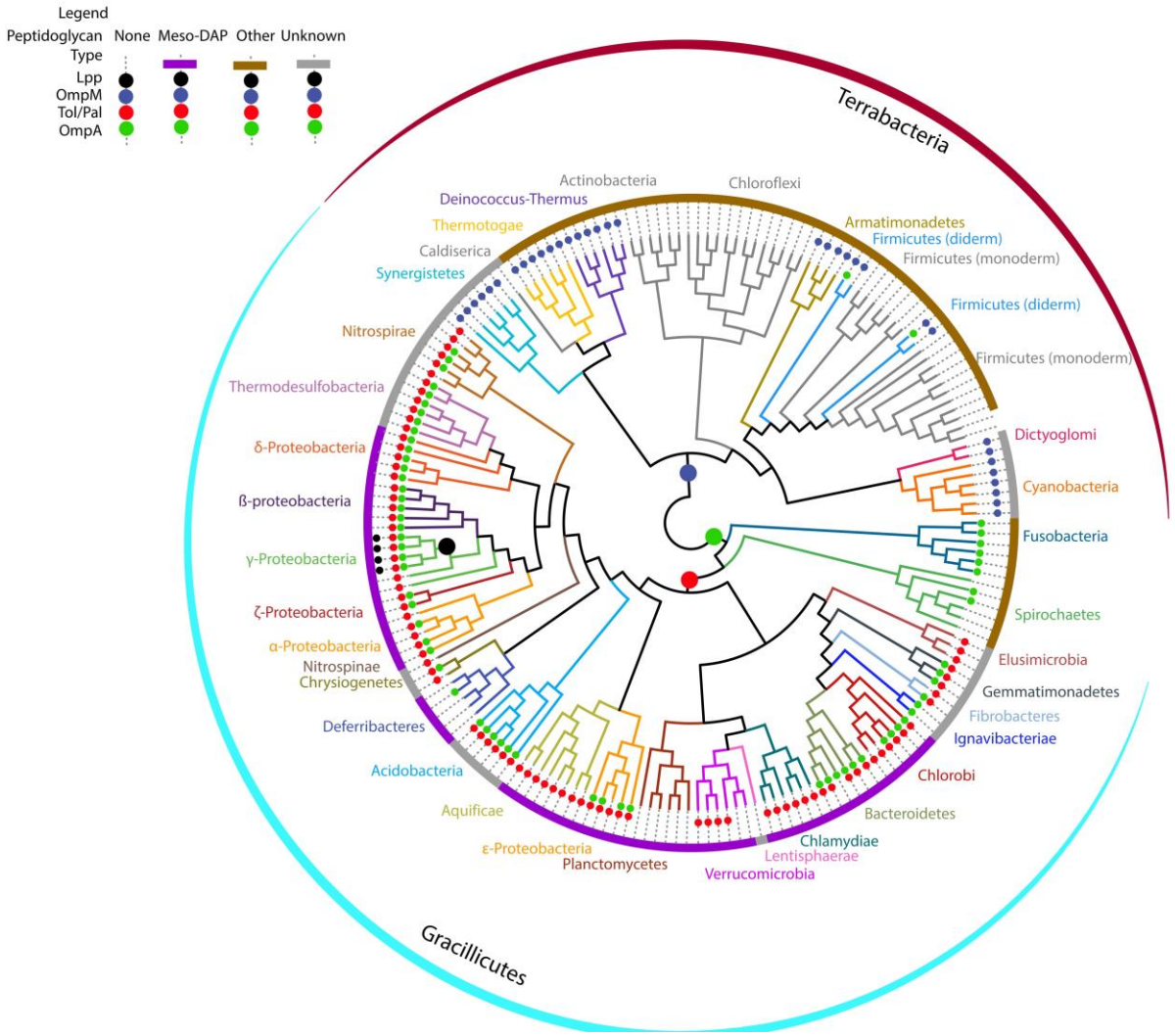

Figure 2: Distribution of the four OM tethering systems in Bacteria.
Schematic tree of Bacteria, based on a concatenated dataset of RNA polymerase ß and ß' subunits (1821 amino acid positions). Filled circles represent the presence of Braun's lipoprotein (Lpp) (dark green), TolB/Pal complex (red), OmpA (light green), and OmpM (blue). Solid bars represent PG.

RESULTS AND DISCUSSION

We plotted the presence/absence of homologues of Lpp, TolB/Pal, OmpA, and OmpM on a phylogeny of Bacteria (Figure 2, Table S1, and M&M).

The monoderm bacterial phyla, including Chloroflexi, monoderm Firmicutes, and Actinobacteria possess no homologs of these proteins, providing an ideal negative control (in grey in Figure 2, Table S1). Caldiserica (former candidate phylum OP5) is a recently described phylum which was portrayed as diderm due to the presence of 3-hydroxy fatty acid because it is frequently found attached to LPS (Mori *et al.*, 2009). However, this molecule is likely present for another purpose, or the Caldiserica have a novel type of OM, as we could not detect any of the known key OM markers such as enzymes of LPS biosynthesis or β-barrel synthesis (data not shown). Mycobacteria also lack any form of known OM attachment (Table S1) although they have an atypical OM that is directly bound to PG layer. In these bacteria, the mycolic acids, which constitute the OM, are bound to arabinogalactan which in turn are bound to the PG. These components form a single molecule, avoiding the need for a protein-based attachment mechanism (Bansal-Mutalik and Nikaido, 2014).

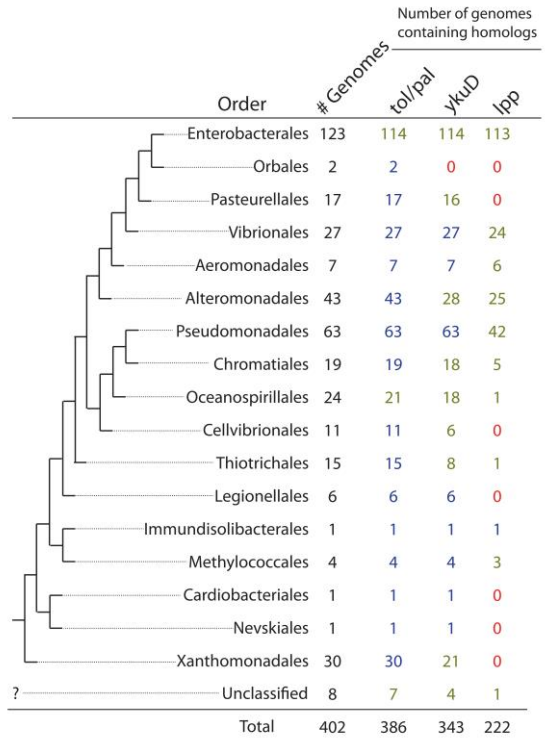

Figure 3: Distribution of Lpp in γ-Proteobacteria. Phylogeny is based on the literature (Campbell et al., 2014). The first column contains the total number of genomes in black, the second contains the distribution of TolB/Pal systems, the third contains the distribution of YkuD, which is responsible for Lpp attachment to PG, the fourth column represents the distribution of Lpp

The archetypal OM attachment mechanism, Lpp, has the lowest distribution among bacterial phyla, being patchily present in γ-Proteobacteria only (Figure 2 and Table S1). In the absence of Lpp, one of the alternative mechanisms for OM-tethering must be in place. To further explore its distribution in this bacterial group, we proceeded to perform an exhaustive search in a local databank including 402 genomes corresponding to all currently described species of γ-Proteobacteria with complete genomes (M&M, Figure 3, and Table S2). Lpp appears to not be universally distributed among the γ-Proteobacteria and is completely absent from most basal groups. Generally, we observe co-presence of Lpp with YkuD, the domain responsible for the covalent bond. However, there are some instances where Lpp exists without YkuD. This suggests the existence of a possible alternative enzyme for the covalent bond in these lineages, or that Lpp has another function than OM attachment (e.g. just being surface exposed). In contrast, TolB/Pal is much more widely spread and might be the main OM-attachment system in the γ-Proteobacteria. Finally, some genomes lack Lpp and TolB/Pal altogether, and it would be interesting to investigate experimentally what is used to attach the OM.

Consistently with these results, the distribution of TolB/Pal among all bacterial phyla is widespread (Figure 2). Surprisingly, it appears restricted to the Gracilicutes (Figure 2). In contrasts, OmpM is found specifically in the diderm Terrabacteria (Battistuzzi and Hedges, 2009) (Figure 2). Concerning OmpA, its presence is non-uniformly distributed and most of our analysed genomes do not possess any homologue, in particular among the diderm Terrabacteria, apart from two diderm Firmicutes (Figure 2 and Table S1). Such bimodal distribution of OM-tethering mechanisms which predominantly match the phylogeny of Bacteria is stunning and might reflect the evolution of these systems during the divergence of the major phyla. We propose the following evolutionary scenario (Figure 4): OmpM-based OM attachment is the most ancient system that appeared very early on in bacterial evolution and represented the main mechanism of all diderm Terrabacteria; this attachment mechanism was lost at the divergence of Fusobacteria and coincided with the origin of OmpA; Fusobacteria and Spirochaetes would represent a sort of "transitional" state, while the TolB/Pal system appeared later in the Gracilicutes (Cavalier-Smith, 2006b) and gradually replaced OmpA, becoming the main OM-attachment system in this clade; finally Lpp represents a recent addition to the TolB/Pal system in γ-Proteobacteria. Therefore, the fact of possessing one system of the other might represent an interesting marker for the phylogeny of Bacteria, and the potential placement of newly sequenced diderm phyla of unclear affiliation.

The reasons for these important transitions in OM-attachment mechanisms in Bacteria (from OmpM to OmpA to TolB/Pal, to Lpp) are puzzling and worth of speculation. They might reside in the different ways these systems attach to the PG. We, therefore, carried out an exhaustive survey of the available literature for PG types and mapped this information on the bacterial tree (Figure 2). Although most lineages remain totally uncharacterized, it appears that the presence of Meso-DAP follows the distribution of TolB/Pal systems in the Gracilicutes and the use of D-OmpA for binding, while the Terrabacteria, as well as the 'transitional' Fusobacteria and Spirochaetes, seem more variable in their PG types (Figure 2).

The binding of PG by OmpM in members of the Negativicutes has been shown to occur through a polyamine modification of the PG peptide chain (Kamio and Nakamura, 1987). This modification has not been demonstrated in the other diderm Terrabacteria clades,

but might also be present. This mid synthesis addition to the PG is likely unnecessary in the Thermotogae as some incorporate lysine into the PG side-chain and therefore the amine group is exposed for binding (Boniface *et al.*, 2009). Deinococcus-Thermus may use a similar mechanism except with ornithine replacing lysine (Quintela *et al.*, 1999). Unfortunately, early studies on Cyanobacteria were biased against atypical amino acids of the PG by removing them with digestion (Jürgens *et al.*, 1983), however subsequent work has shown the presence of lysine in *Anabena cylindrica* (Kodani *et al.*, 1999). We anticipate that Armatimonadetes, Synergistetes, and Dictyoglomus also have a similarly exposed amine present on their PG, which will need to be proven experimentally. This would represent a clear physio-chemical reason for the incorporation of amine side-chains within the PG.

The Spirochaetes are a peculiar case, as a few have OmpA homologues, and some have no attachment mechanism at all (Figure 2 and Table S1). This phylum possesses many odd and unique features in their cell envelope including the absence of LPS in some members, flagella inserted into the periplasmic space, and surface display mechanisms. No definitive studies have been done and while some Spirochaetes possess OmpA like protein, it has been suggested that the function of OM attachment within this group is done by D-OmpA lipoproteins (found in abundance) including the Leptospira pathogenicity factor Loa22 (Koizumi and Watanabe, 2003). Spirochaetes make strong use of lipoproteins in the OM and it would be logical to attach via a lipoprotein domain as opposed to a β-barrel.

Members of the PVC superphylum have very peculiar cell envelopes but have not been studied extensively. Within this group a conserved pair of proteins is found which may function like the leptospiral lipoproteins. These two proteins lie side-by-side in the genomes and are both homologs of Pal, including a lipoprotein signal and D-OmpA. This pair is likely derived from the Tol-Pal system and, due to the unique envelope structure, requires a different system (Fuerst, 2013).

Little is known about the unique thermophilic Deferribacteres. Their OM has had very little study and our analysis did not discover any Tol-Pal system, however each genome had many D-OmpA containing proteins, including one with a predicted β-barrel which may function in attachment but experimental work is necessary.

It is possible that the Tol-Pal system emerged initially to perform a simple function, OM attachment, and that originally consisted of only Pal and the β-propeller domain of TolB. This complex was inserted within the OM via lipidation and tightly clamped the PG between these two proteins. In contrast, Lpp would have emerged in the γ-Proteobacteria, probably to provide what is the strongest and least metabolically costly protein, as it covalently binds to the PG and integrates into the OM via a lipid. As many Lpp-containing γ-Proteobacteria are pathogens, one advantage to replace a β-barrel by Lpp for OM attachment might lie in a better shielding from the immune system of the host, but this hypothesis remains totally speculative at this stage.

In conclusion, our results illustrate a large diversity of OM-tethering systems in all major bacterial diderm phyla. Not only do they allow to discuss important evolutionary hypotheses, but also open the way to understand the nature of the very first bacterial ancestor. Finally, they pave the way to further study of the vast diversity of uncharacterized bacterial cell envelopes.

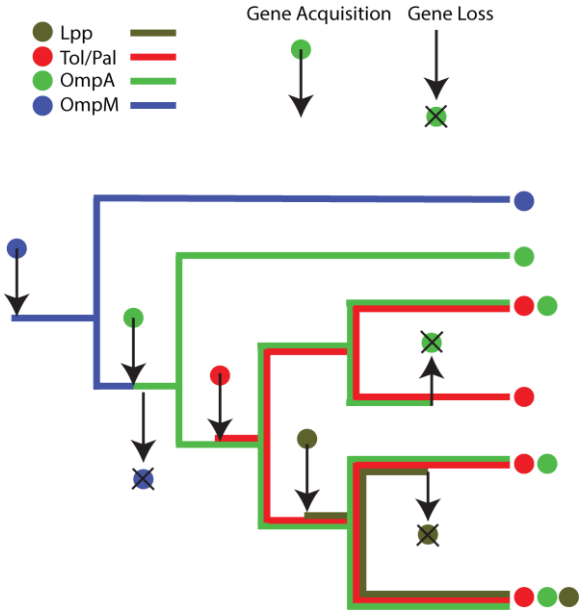

Figure 4: Evolutionary Hypothesis for OM Attachment.

MATERIALS AND METHODS

DATABASES AND HMM SEARCH

We assembled a local databank of 198 genomes from a wide representative sampling of the major bacterial phyla that have complete genome sequences (Supplementary file 1). Exhaustive HMM-based homology searches were carried out on this genome databank by using the HMMER package (Mistry *et al.*, 2013) with HMM profiles of each protein from the PFAM 29.0 database (Finn *et al.*, 2015, http://pfam.xfam.org) as a query. Additional searches with tBLASTn (Camacho *et al.*, 2009) were used to identify eventually misannotated homologues in some genomes.

The database of γ-Proteobacteria was constructed by gathering all complete genomes from NCBI and selecting named species, leading to 402 genomes.

TREE CONSTRUCTION

Alpha and beta subunits of RNA polymerase were aligned with MAFFT v7.055 (Yamada *et al.*, 2016) with default parameters, and unambiguously aligned positions were selected with BMGE 1.1 (Criscuolo and Gribaldo, 2010) and the BLOSUM30 substitution matrix. Trimmed datasets were concatenated and PhyloBayes v3.3b (Lartillot et al., 2009) was used to perform Bayesian analysis using the evolutionary model CAT+GTR+Γ4. Two independent chains were run until convergence, assessed by evaluating the discrepancy of bi-partition frequencies between independent runs. The first 25% of trees were discarded as burn-in and the posterior consensus was computed by selecting one tree out of every two.

LPP DETECTION

Due to its small size, a number of methods were used to search for Lpp homologues. Experimentally proven Lpp homologs were initially used to identify clear homologues in a selection of representative bacteria. Results were checked by analysis of multiple alignments with MAFFT (Yamada *et al.*, 2016), and genome synteny with SyntTax (Oberto, 2013). An HMM model was then built with these sequences and used to query the whole database. We selected sequences that had at least 70% coverage with the HMM model and E-value less than 0.00001. We proceeded by iteration until more significant results could not be found. YkuD homologues were also searched for the more detailed analysis of Lpp distribution in the γ-Proteobacteria.

OMPA DETECTION

OmpA was defined for the scope of this study as a protein containing a D-OmpA (PF00691) and a β-barrel. An initial search for D-OmpA was conducted and these results were limited to those with a positive β-barrel prediction by PRED-TMBB (Bagos *et al.*, 2004). Each positive β-barrel prediction was further checked using BOCTOPUS2 (Hayat *et al.*, 2016) until a positive result was found. If no protein was found with a positive detection of D-OmpA, PRED-TMBB, and BOCTOPUS2 it was marked as absent from the genome.

TOLB/PAL DETECTION

In order to distinguish it from other proteins possessing a similar domain, Pal was defined for the scope of this study as a protein containing an N-terminal D-OmpA domain (PF00691) and having a TolB homologue in the immediate genetic area. TolB proteins were required to have a beta propeller repeat (PF07676 or PF00400). We did not consider the inner membrane ATPase complex as this is not universally distributed in Tol-Pal systems (Sturgis, 2001).

OMPM DETECTION

OmpM was defined for the scope of this study as a protein containing an N-terminal SLH domain (PF00691) and a β-barrel. The same method used for OmpA was utilized here. An exemption was made for the Thermotogales, which have a two protein-system with an SLH domain-containing protein (Ompα) immediately upstream of a gene containing a β-barrel (Omp-β). No such architectures were detected outside the Thermotogales.

REFERENCES

Adu-Bobie, J., Lupetti, P., Brunelli, B., Granoff, D., Norais, N., Ferrari, G., et al. (2004) GNA33 of Neisseria meningitidis is a lipoprotein required for cell separation, membrane architecture, and virulence. Infect Immun 72: 1914–9

Antunes, L.C., Poppleton, D., Klingl, A., Dupuy, B., Brochier-Armanetd, C., Beloin, C., and Gribaldo, S. (2016) Phylogenomic analysis supports the ancestral presence of LPS outer membranes in the Firmicutes. Elife e14589

Bagos, P.G., Liakopoulos, T.D., Spyropoulos, I.C., and Hamodrakas, S.J. (2004) PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. Nucleic Acids Res 32: W400-4

Bansal-Mutalik, R., and Nikaido, H. (2014) Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. Proc Natl Acad Sci U S A 111: 4958–63

Battistuzzi, F.U., and Hedges, S.B. (2009) A major clade of prokaryotes with ancient adaptations to life on land. Mol Biol Evol 26: 335–43

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235–42

Boniface, A., Parquet, C., Arthur, M., Mengin-Lecreulx, D., and Blanot, D. (2009) The elucidation of the structure of Thermotoga maritima peptidoglycan reveals two novel types of cross-link. J Biol Chem 284: 21856–62

Braun, V., and Rehn, K. (1969) Chemical characterization, spatial distribution and function of a lipoprotein (murein-lipoprotein) of the E. coli cell wall. The specific effect of trypsin on the membrane structure. Eur J Biochem 10: 426–38.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. BMC Bioinformatics 10: 421

Campbell, C., Sutcliffe, I.C., and Gupta, R.S. (2014) Comparative proteome analysis of Acidaminococcus intestini supports a relationship between outer membrane biogenesis in Negativicutes and Proteobacteria. Arch Microbiol 196: 307–310

Cascales, E., Bernadac, A., Gavioli, M., Lazzaroni, J.-C., and Lloubes, R. (2002) Pal lipoprotein of Escherichia coli plays a major role in outer membrane integrity. J Bacteriol 184: 754–9

Cavalier-Smith, T. (2006) Cell evolution and Earth history: stasis and revolution. Philos Trans R Soc Lond B Biol Sci 361: 969–1006

Chen, Z.-Y., Cao, J., Xie, L., Li, X.-F., Yu, Z.-H., and Tong, W.-Y. (2014) Construction of leaky strains and extracellular production of exogenous proteins in recombinant Escherichia coli. Microb Biotechnol 7: 360–70

Ching, G., and Inouye, M. (1985) Evolution of the lipoprotein gene in the enterobacteriaceae. Cloning and DNA sequence of the lpp gene from Proteus mirabilis. J Mol Biol 185: 501–7

Confer, A.W., and Ayalew, S. (2013) The OmpA family of proteins: Roles in bacterial pathogenesis and immunity. Vet Microbiol 163: 207–222.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., et al. (2015) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44: D279–D285

Fuerst, J.A. (2013) Planctomycetes : cell structure, origins and biology. Humana Press, .

Godlewska, R., Wiśniewska, K., Pietras, Z., and Jagusztyn-Krynicka, E.K. (2009) Peptidoglycan-associated lipoprotein (Pal) of Gram-negative bacteria: function, structure, role in pathogenesis and potential application in immunoprophylaxis. FEMS Microbiol Lett 298: 1–11

Guo, M.S., Updegrove, T.B., Gogol, E.B., Shabalina, S.A., Gross, C.A., and Storz, G. (2014) MicL, a new σE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. Genes Dev 28: 1620–34

Hansel, A., Pattus, F., Jürgens, U.J., and Tadros, M.H. (1998) Cloning and characterization of the genes coding for two porins in the unicellular cyanobacterium Synechococcus PCC 6301. Biochim Biophys Acta - Gene Struct Expr 1399: 31–39.
Hayat, S., Peters, C., Shu, N., Tsirigos, K.D., and Elofsson, A. (2016) Inclusion of dyad-repeat pattern improves topology prediction of transmembrane β-barrel proteins. Bioinformatics 32: 1571–3
Jürgens, U.J., Drews, G., and Weckesser, J. (1983) Primary structure of the peptidoglycan from the unicellular cyanobacterium Synechocystis sp. strain PCC 6714. J Bacteriol 154: 471–8
Kamio, Y., and Nakamura, K. (1987) Putrescine and cadaverine are constituents of peptidoglycan in Veillonella alcalescens and Veillonella parvula. J Bacteriol 169: 2881–4 h
Kim, Y.C., Tarr, A.W., and Penfold, C.N. (2014) Colicin import into E. coli cells: A model system for insights into the import mechanisms of bacteriocins. Biochim Biophys Acta - Mol Cell Res 1843: 1717–1731
Kodani, S., Ishida, K., and Murakami, M. (1999) Occurrence and identification of UDP-N-acetylmuramyl-pentapeptide from the cyanobacterium Anabaena cylindrica. FEMS Microbiol Lett 176.
Koizumi, N., and Watanabe, H. (2003) Molecular cloning and characterization of a novel leptospiral lipoprotein with OmpA domain. FEMS Microbiol Lett 226: 215–9
Kojima, S., and Kamio, Y. (2012) Molecular basis for the maintenance of envelope integrity in Selenomonas ruminantium: cadaverine biosynthesis and covalent modification into the peptidoglycan play a major role. J Nutr Sci Vitaminol (Tokyo) 58: 153–60
Kojima, S., Kaneko, J., Abe, N., Takatsuka, Y., and Kamio, Y. (2011) Cadaverine covalently linked to the peptidoglycan serves as the correct constituent for the anchoring mechanism between the outer membrane and peptidoglycan in Selenomonas ruminantium. J Bacteriol 193: 2347–2350.
Llamas, M.A., Ramos, J.L., and Rodríguez-Herva, J.J. (2000) Mutations in each of the tol genes of Pseudomonas putida reveal that they are critical for maintenance of outer membrane stability. J Bacteriol 182: 4764–72
Lupas, A., Müller, S., Goldie, K., Engel, A.M., Engel, A., and Baumeister, W. (1995) Model Structure of the Ompα Rod, a Parallel Four-stranded Coiled Coil from the HyperthermophilicEubacteriumThermotoga maritima. J Mol Biol 248: 180–189.
Magnet, S., Dubost, L., Marie, A., Arthur, M., and Gutmann, L. (2008) Identification of the L,D-transpeptidases for peptidoglycan cross-linking in Escherichia coli. J Bacteriol 190: 4782–5
Misra, C.S., Basu, B., and Apte, S.K. (2015) Surface (S)-layer proteins of Deinococcus radiodurans and their utility as vehicles for surface localization of functional proteins. Biochim Biophys Acta - Biomembr 1848: 3181–3187.
Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., and Punta, M. (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res 41: e121
Mori, K., Kaoru, Y., Yayoi, S., Urabe, T., and Suzuki, K. (2009) Caldisericum exile gen. nov., sp. nov., an anaerobic, thermophilic, filamentous bacterium of a novel bacterial phylum, Caldiserica phyl. nov., originally called the candidate phylum OP5, and description of Caldisericaceae fam. nov., Caldisericales ord. no. Int J Syst Evol Microbiol 59: 2894–8.
Oberto, J. (2013) SyntTax: a web server linking synteny to prokaryotic taxonomy. BMC Bioinformatics 14: 4
Park, J.S., Lee, W.C., Yeo, K.J., Ryu, K.-S., Kumarasiri, M., Hesek, D., et al. (2012) Mechanism of anchoring of OmpA protein to the cell wall peptidoglycan of the gram-negative bacterial outer membrane. FASEB J 26: 219–28
Petrus, A.K., Swithers, K.S., Ranjit, C., Wu, S., Brewer, H.M., Gogarten, J.P., et al. (2012) Genes for the major structural components of Thermotogales species' togas revealed by proteomic and evolutionary analyses of OmpA and OmpB homologs. PLoS One 7: e40236
Poppleton, D.I., Duchateau, M., Hourdel, V., Matondo, M., Flechsler, J., Klingl, A., et al. (2017) Outer Membrane Proteome of Veillonella parvula: A Diderm Firmicute of the Human Microbiome. Front Microbiol 8: 1215
Quintela, J.C., García-del Portillo, F., Pittenauer, E., Allmaier, G., and Pedro, M.A. de (1999) Peptidoglycan fine structure of the radiotolerant bacterium Deinococcus radiodurans Sark. J Bacteriol 181: 334–7
Schwechheimer, C., Kulp, A., Kuehn, M.J., Berleman, J., Auer, M., Deatherage, B., et al. (2014) Modulation of bacterial outer membrane vesicle production by envelope structure and content. BMC Microbiol 14: 324
Shrivastava, R., Jiang, X., and Chng, S.S. (2017) Outer membrane lipid homeostasis via retrograde phospholipid transport in Escherichia coli. bioRxiv
Shu, W., Liu, J., Ji, H., and Lu, M. (2000) Core structure of the outer membrane lipoprotein from Escherichia coli at 1.9 A resolution. J Mol Biol 299: 1101–12
Sturgis, J.N. (2001) Organisation and evolution of the tol-pal gene cluster. J Mol Microbiol Biotechnol 3: 113–22
Sugawara, E., Nestorovich, E.M., Bezrukov, S.M., and Nikaido, H. (2006) Pseudomonas aeruginosa porin OprF exists in two different conformations. J Biol Chem 281: 16220–9
Wessel, A.K., Liew, J., Kwon, T., Marcotte, E.M., and Whiteley, M. (2013) Role of Pseudomonas aeruginosa peptidoglycan-associated outer membrane proteins in vesicle formation. J Bacteriol 195: 213–9
Whitman, W.B. (2011) Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes. 2nd Editio., Springer, .
Yamada, K.D., Tomii, K., and Katoh, K. (2016) Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees. Bioinformatics btw412

3 Find a mechanism of transition between diderm and monoderm Firmicutes.

The third and final goal of my thesis was to determine what evolutionary mechanism resulted in the monoderm/diderm transition. I chose to explore the peptidoglycan enzymes as they had been proposed by Cavalier-Smith as a mechanism of transition (Cavalier-Smith, 2006a).

3.1 Survey of Peptidoglycan catalysing domains in Firmicutes

At the onset of my thesis, Dr Gribaldo and I were discussing potential mechanisms of the diderm/monoderm transition. At this point, we had not yet firmly established whether the OM was lost in the Firmicutes or transferred. One of the hypotheses that we discussed was through controlling PG growth. Should the diderm first hypothesis prove to be true, a loss of regulation in a key PG enzyme would result in the outgrowth of the peptidoglycan rupturing the OM. Conversely, should a duplication happen so that an additional enzyme or regulatory step is added it may be possible for a monoderm to shrink its PG and obtain an OM through sporulation.

To explore this hypothesis I built upon the work of (Layec *et al.*, 2008). In their work, they conducted a survey of peptidoglycan enzymes in the available Firmicute genomes. They focused on the architecture of these proteins and what domains were on each protein apart from the catalytic domain. Unfortunately, they had no diderm Firmicute genomes included in their dataset.

Expanding on the same methodology that the authors proposed I performed my analysis using 80 genomes and 39 domains. Additionally, I included many diderm Firmicute genomes to enable accurate comparison of monoderm and diderm types. After this survey was complete I found some interesting PG enzymes of the Negativicutes that were not previously described including a FtsW homolog and penicillin binding protein. However, my main result was the identification of *spoIID* in the non-sporulating diderm Firmicutes. This enzyme immediately suggests a mechanism for the diderm/monoderm transition.

This data is not currently planned for publication, as we were unable to complete the follow-up work in mutating *spoIID* (Following Section).

Survey of peptidoglycan catalysing domains in Firmicutes suggests mechanism of diderm/monoderm transition

One of the most essential components of the Bacterial cell envelope is the peptidoglycan. This molecule is responsible for shape, structure, among other uses. Diderm (Gram-negative) bacteria possess thin peptidoglycan, one to two molecules thick, while monoderms (Gram-positive) have peptidoglycan that is several molecules thick. Fortunately, a clade of Bacteria exists that has both types, the Firmicutes. By comparing the 39 peptidoglycan catalytic domains from 80 genomes we were able to draw conclusions on the unique characteristics of Firmicute peptidoglycan. Additionally, we discovered an enzyme that may be responsible for the diderm/monoderm transition within Bacterial

Peptidoglycan | Genomics | Sporulation | Evolution

For an in depth discussion of peptidoglycan structure, synthesis, and remodelling, please see Introduction Section 1.4 of this dissertation.

The peptidoglycan of Firmicutes is of the most diverse and well-studied apart from *E. coli* (Vollmer *et al.*, 2008). This peptidoglycan has been divided into many different types based on the branching pattern which can change from strain to strain in monoderms. Conversely, the PG of diderm Firmicutes is much simpler, consisting of a meso-DAP linkage with sporadic polyamine modifications throughout the genome (Schleifer and Kandler, 1972; Kamio and Nakamura, 1987).

Two previous studies have used bioinformatics methods for broad PG remodelling enzymes: one of the Firmicutes (Layec *et al.*, 2008) and one of the mycobacteria(Machowski *et al.*, 2014). The mycobacterial study focused on resuscitation promoting factors (Rpfs), penicillin binding proteins, endopeptidases, L, D-transpeptidases and N-acetylmuramoyl-L-alanine amidases. They found at least one copy of each of these enzymes and that they were highly conserved among the Mycobacteria.

More relevant to this study is the work of Layec et al., which used PFAM searches to look for PG enzymes. They focused on amidases, peptidases, and lytic transglycosylases. The authors searched for these genes in 133 genomes belonging to 22 genera. They looked at accessory domains of the PG hydrolases and whether they were phage-encoded. The final focus of the paper was to find the enzymes responsible for daughter cell-separation. They found that these enzymes were species-specific, but did not go into further depth on any of the other enzymes they studied.

In this study, I compared the enzymes of PG from monoderm Firmicutes with the diderm Firmicutes using an *in silico* approach. I took all known PFAM domains corresponding to catalysis of PG and searched a representative selection of Firmicutes for the presence of the domains. This will build on the previous study by Layec and provide a more in depth analysis including diderm Firmicutes. From this data, I hoped to understand why a drastic difference in PG depth is found between monoderm and diderm bacteria. This data will increase our understanding of PG in evolution, function, and explore the PG in an unstudied clade, the Negativicutes.

Significance

Peptidoglycan is a fundamental molecule of Bacterial physiology, which has been a priority antibiotic target since their advent. Understanding the distribution of catatalytic sites enables greater discovery potential for antibiotics. The evolutionary question of diderm/monoderm transition may prove to be one of the greatest questions in bacterial physiology.

Results

We studied seventy-two genomes representing members from nine orders including two diderm classes, the Negativicutes and the Halanaerobiales, one order without PG, the Mollicutes, and six orders with monoderm architecture, the Natranaerobiales, Bacillales, Lactobacillales, Clostridia Thermoanaerobacteriales, and the Erysipelotrichia. The genomes were searched for the presence of enzymes or domains associated with PG biosynthesis or remodelling.

Ribosomal Reference Tree. To compare our data and search results we first constructed a robust phylogenetic tree (Figure 1) from a concatenation of ribosomal proteins. It has discrepancies from the best current phylogeny (Antunes *et al.*, 2016), as the Clostridia are not monophyletic. This discrepancy is minor as the rest of the tree is congruent with the data and better informs the phylogeny of the genomes chosen.

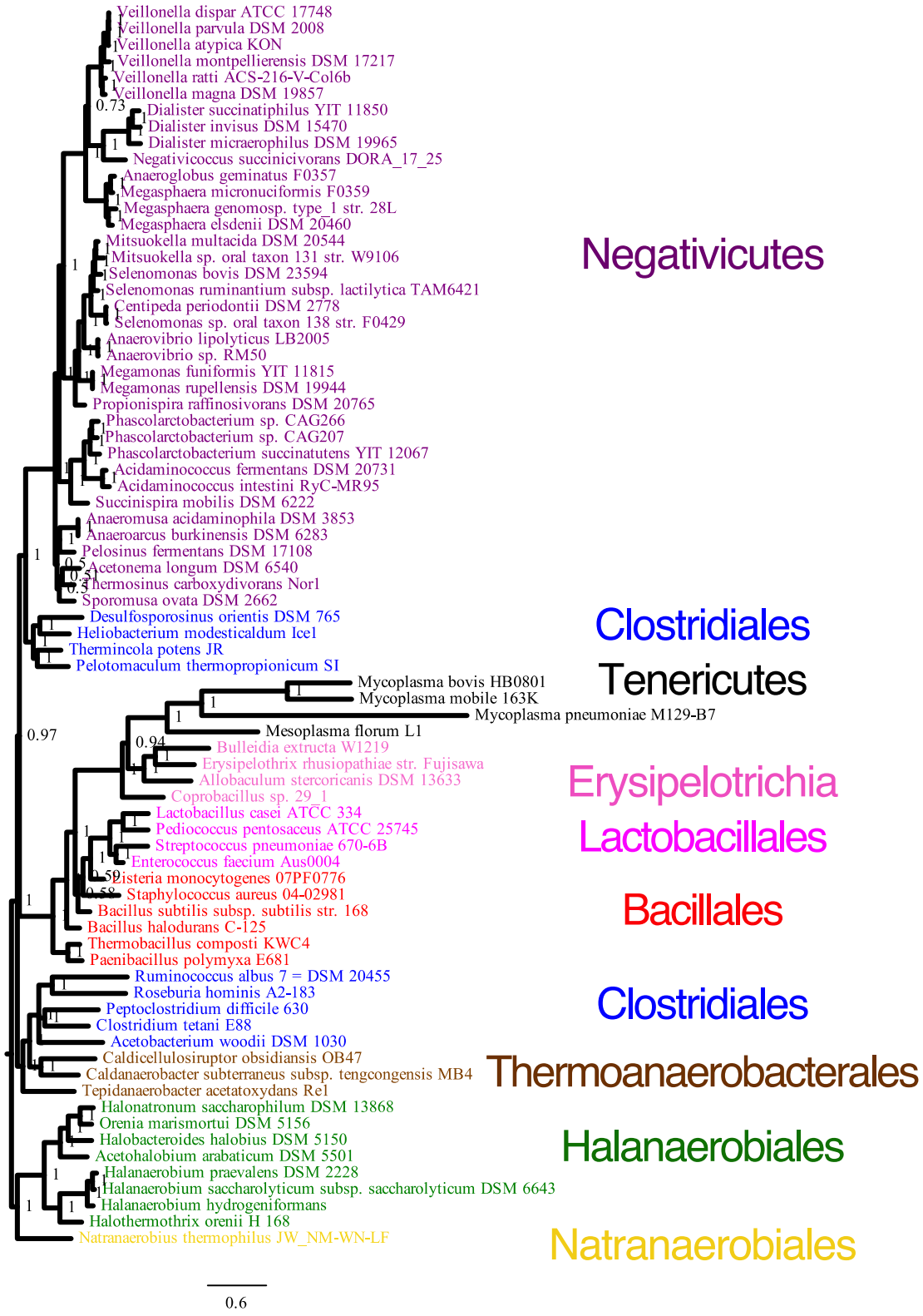

Fig 1. Ribosomal Reference Tree: Phylogeny of the Firmicutes derived from ribosomal proteins. The tree was constructed using PhyloBayes with 6206 residues using GTR+CAT model of substitution.

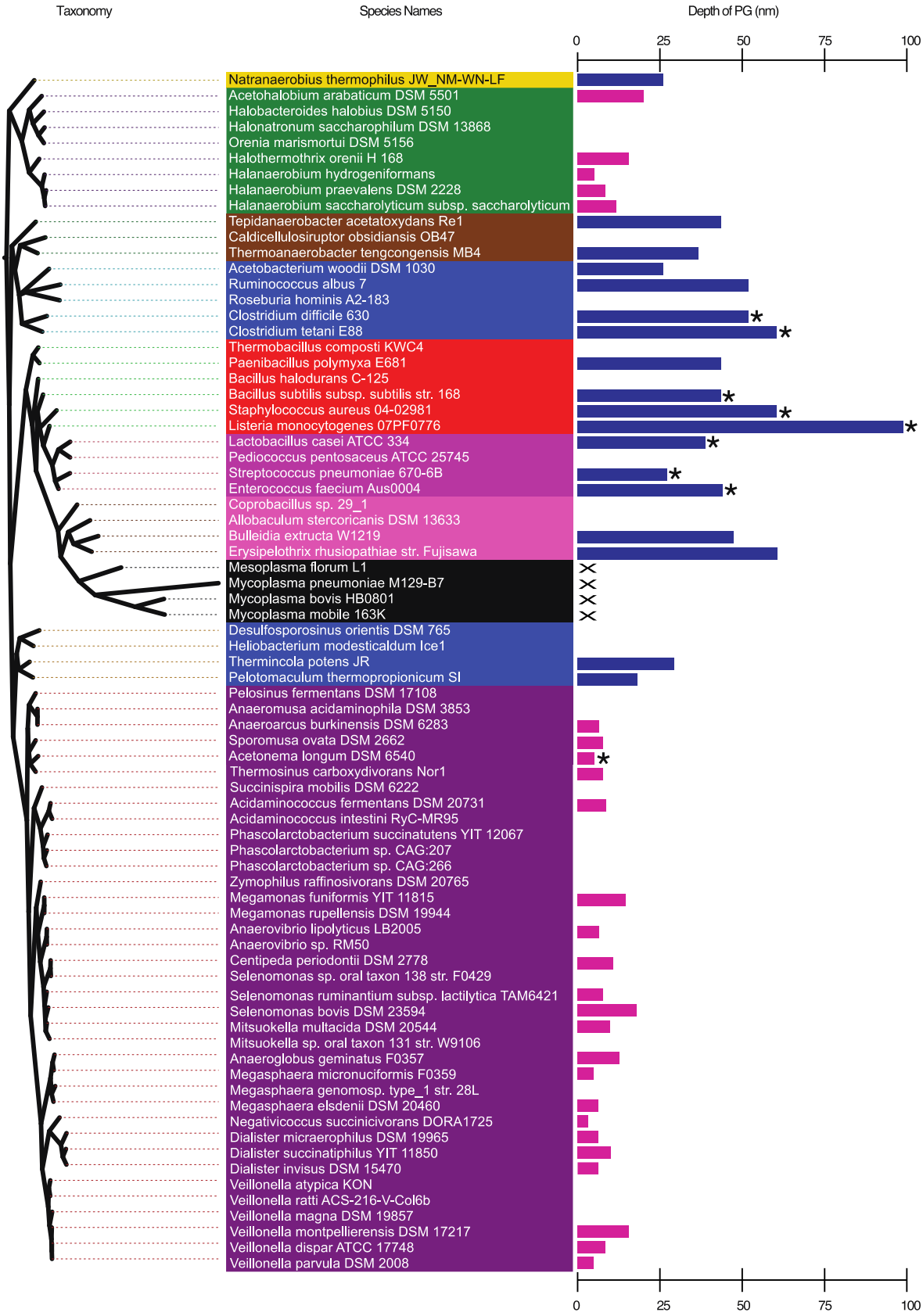

Fig 2. The thickness of PG in selected bacteria. Published experimental data for PG thickness in genomes studied. Pink bars represent diderm organisms, while dark blue represent monoderms. Background colours correspond to individual orders. * Denote data obtained from cryo-EM. "X" denotes complete absence of peptidoglycan

Peptidoglycan thickness. To determine if our question was valid, we first compiled known data on PG thickness for our organisms of study (Fig 2). Although little quality cryo-EM work has been done, the thickness in diderms is generally between 5-15nm, while monoderms tend to possess PG 20-80nm wide. This drastic change validates our study and truly proposes the question "Why is the PG thicker in monoderms?"

Core cytoplasmic peptidoglycan biosynthetic enzymes. The core cytoplasmic enzymes of PG biosynthesis, MurA-G and ddl, were found in all genomes studied except the Mollicutes, as expected (Fig 3, Supplementary Table 1). MurA is the first committed step of PG biosynthesis and functions by adding a PEP molecule to UDP-GlcNAc. It was present in multiple copies in many of the genomes and ranged from one to three without any clear distinction by class. The second enzyme in the pathway, MurB, is responsible for reducing the molecule as to prepare it for the addition of peptides. It was present in a single copy except in a few Negativicutes, Halanaerobiales, and Thermoanaerobacteriales. In the next steps of peptidoglycan biosynthesis, a peptide chain is formed by the sequential addition of amino acids to the sugar. The first two enzymes, MurC and MurD, were present in a single copy each. The third enzyme, MurE, was found in two copies in several genomes. The final two amino acids are incorporated as a dimer of D-Alanine. The dimerization enzyme, ddl, was found in one to two copies in all genomes. The addition enzyme, MurF, which adds the dimer, was found as a single copy in all bacteria studied except the Peptococcaceae, which had two copies.

Core Peptidoglycan biosynthetic enzymes involved in lipid-linked intermediates. Before the subunits can be incorporated into the PG sacculus they must first cross the inner membrane. The first step is the attachment of a lipid moiety to the subunit produced by MurF. This step is performed by MurG at the inner membrane to produce a lipid-linked intermediate named lipid I. We found a single copy of MurG present in all PG producing bacteria studied. Lipid I is transferred to MraY where a second sugar moiety is added, GlcNAC, forming lipid II. Again we found a single copy of MraY in all PG producing organisms. The final step is to flip lipid II through the inner membrane to the outer leaflet. Two enzymes have been proposed to do this function in *E. coli*. MurJ (MviN) was absent from most of the diderm bacteria, and Clostridia and absent from all of the Bacilli. The other candidate, FtsW, has two close homologs; RodA is necessary for rod shape and SpoVE is required in sporulation. We found this class of proteins is present in all of the Firmicutes studied except the Tenericutes. Interestingly, we found an additional homolog of the FtsW family in the Negativicutes. (Fig 4), whose function is currently unknown.

High Molecular Weight Penicillin Binding Proteins After transport through the inner-membrane, lipid II has to polymerize and crosslinked with the mature PG. This process is performed by high molecular weight (HMW) penicillin-binding proteins (PBP). Polymerization is performed by Class A HMW PBPs. Class A PBPs consist of two domains a membrane-linked glycosyltransferase domain and a transpeptidase domain. The glycosyltransferase domain performs the polymerization of the chain, while the transpeptidase creates crosslinks with the existing PG structure. The chain is then further crosslinked by Class B HMW PBPs, which do not have a glycosyltransferase domain, instead, they have a dimerization domain used in interactions.

We found the Class A and B PBPs present in all the bacteria studied except the Mollicutes (Fig 5). The class B proteins were present in two or more copies, except for a few non-spore forming Clostridia. The Bacilli, Thermoanaerobacteriales, and Natranaerobiales had a large number of homologs within their genomes. Class A PBPs were usually detected as a single copy, except for the same three clades as the Class B PBPs. Aside from the traditional PBPs we surprisingly found a large number of monofunctional glycosyltransferases (MGT). These proteins consist of a single domain homologous to that of the glycosyltransferase in PBP class A. They are known to extend a PG chain without crosslinking it. We discovered this enzyme in almost every Negativicute studied, but only in four other Firmicutes.

To investigate this we constructed a phylogenetic tree using only the glycosyltransferase domain of PBP class A and the MGTs (Figure 6). From this data, it appears the Negativicutes acquired the gene at time of their separation from the other orders, while the other Firmicutes appear to be transfered, as in the case of Tepidanaerobacter, or recent losses of the transpeptidase domain.

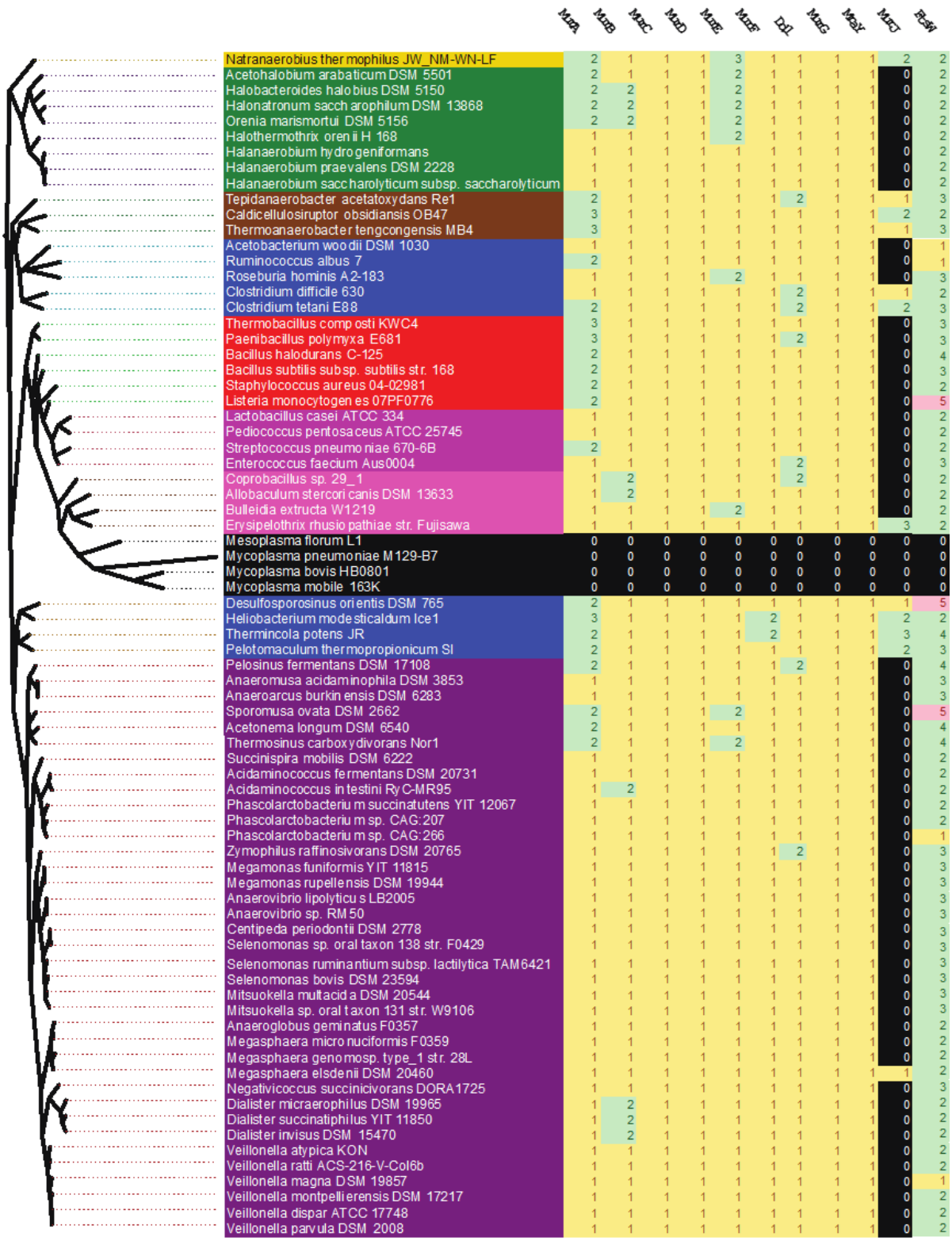

Fig 3. The copy number of Core PG enzymes: Figure shows the total number of homologs found for each enzyme by HMM search. Colours behind numbers indicate copy number, 0=black, 1=yellow, 2-4=green >4=Red. Background colours correspond to individual orders

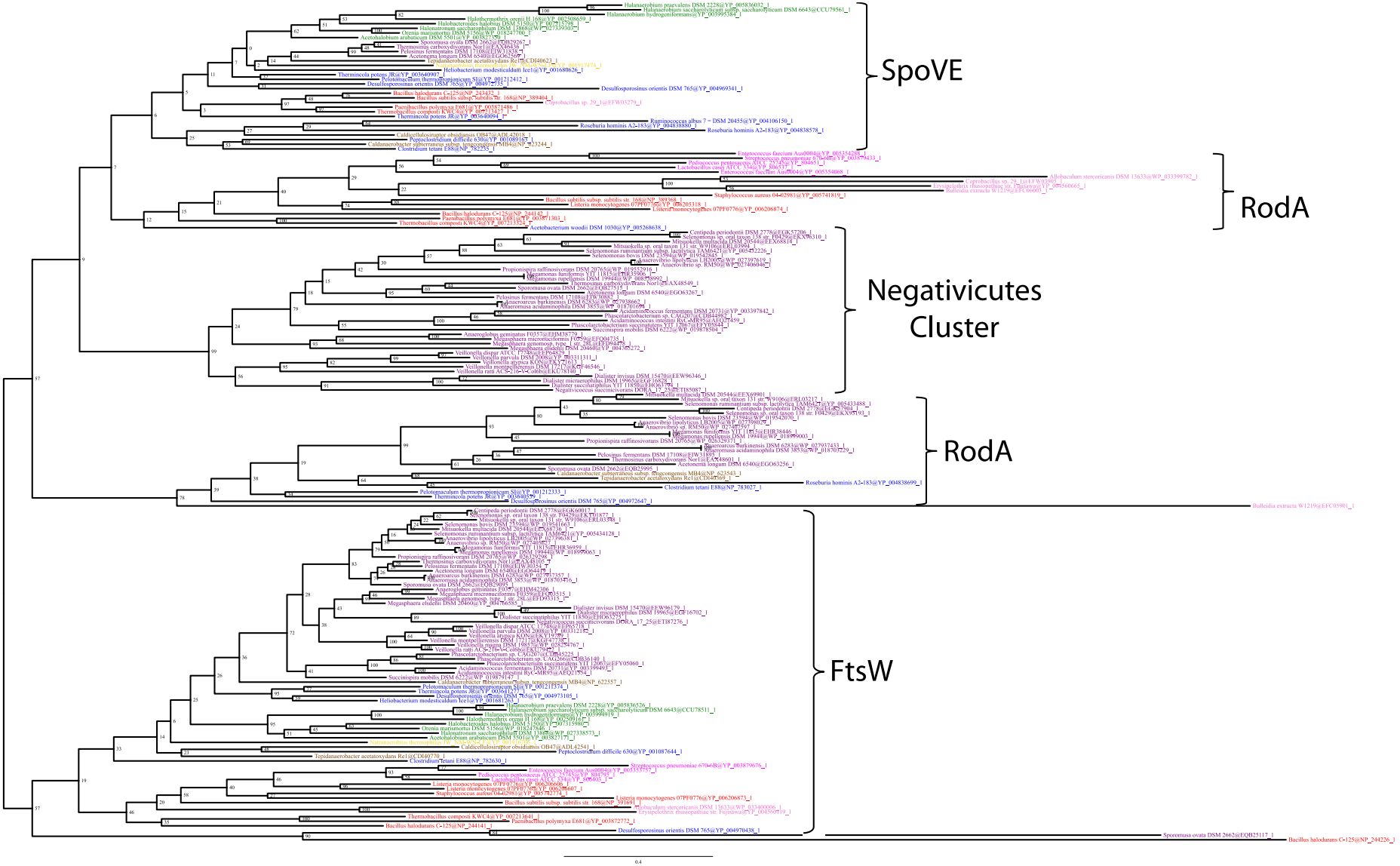

Fig 4. Phylogeny of Firmicute FtsW The tree was constructed with PhyML using 247 residues and an LG model with a Gamma distribution. Protein class determination was determined by a combination of gene annotation, sporulating capability of bacteria, and bacterial shape.

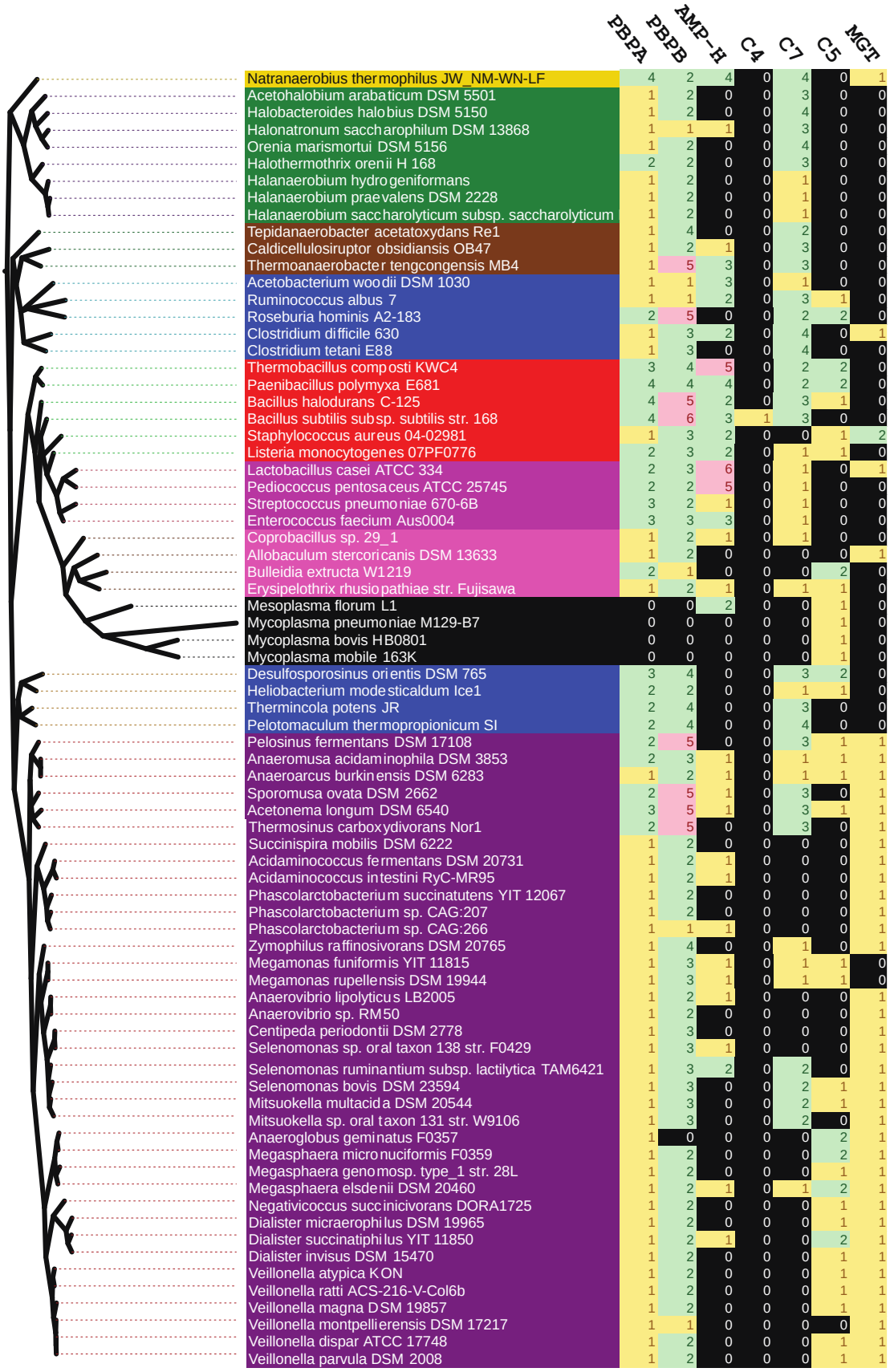

Fig 5. The copy number of Penicillin Binding Proteins: Figure shows the total number of homologs found for each enzyme by HMM search. Colours behind numbers indicate copy number, 0=black, 1=yellow, 2-4=green >4=Red. Background colours correspond to individual orders.

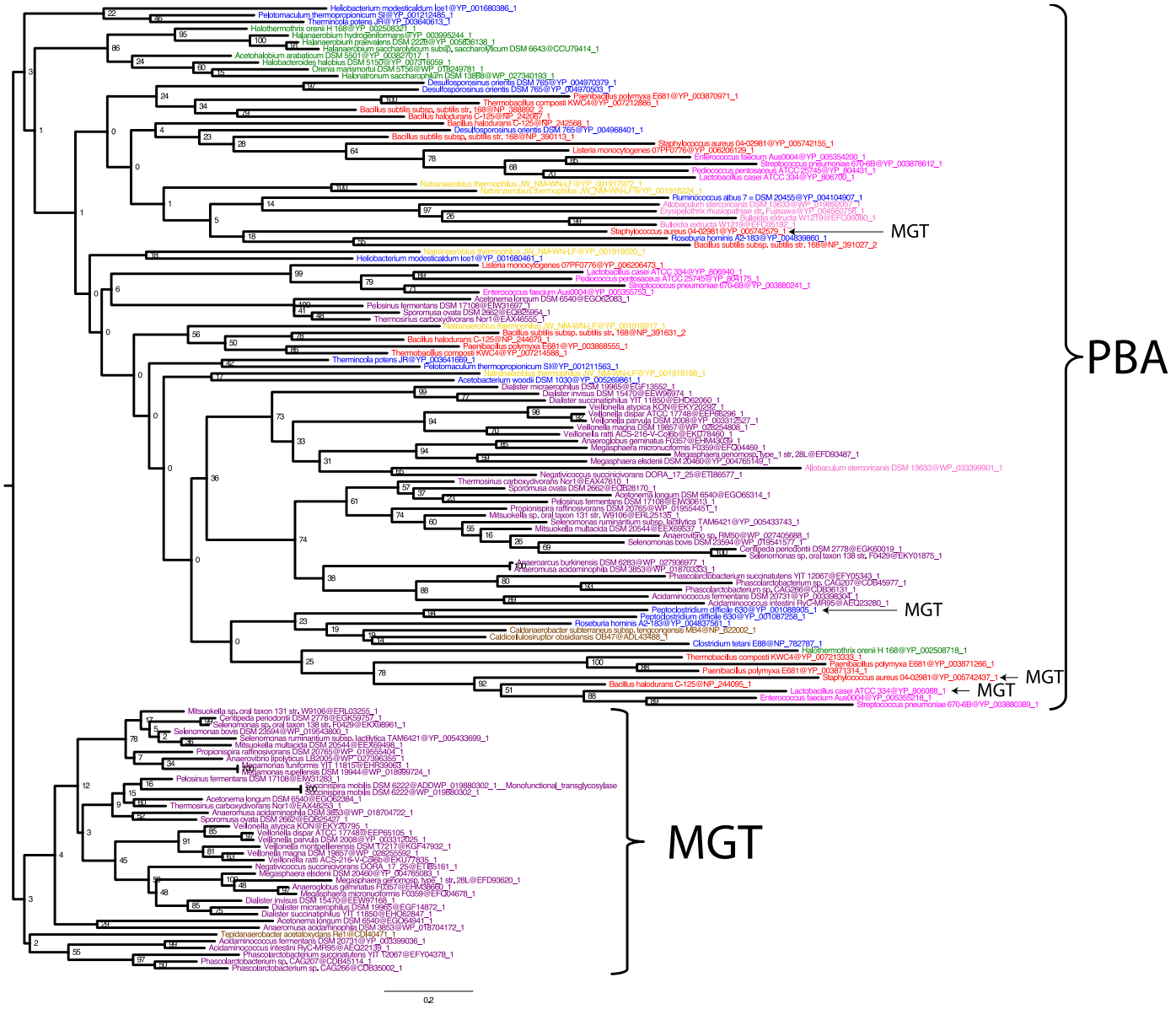

Fig 6. Phylogeny of the Transglycosylase domain of PBPA and MGT: The tree was constructed with PhyML using 166 residues and an LG model with a Gamma distribution.

Division and Cell Wall Cluster.

Recent studies have demonstrated that the enzymes of primary PG biosynthesis form a complex, consisting of MurD, MurE, MurF, MurG, MraW, MraZ, MraY, FtsL, FtsW, PBPA, and FtsZ (Mohammadi *et al.*, 2007)(Favini-Stabile *et al.*, 2013). We investigated the genomic location of the peptidoglycan genes to see if it correlated (Fig 7). We found the cluster of the Negativicutes, Halanaerobiales, Thermoanaerobiales, Natranaerobiales, and Clostridia to have a conserved organization. The scaffold for the protein complex, MurG, was found at the centre of the complex, with MurC-F and MraY as neighbouring genes. The flippase was occasionally found in the centre of the complex, and PBPB was located at the start. The PG genes were flanked by the Fts genes, also called Mre, which are involved in cell division and cytoskeleton structure. No biochemical evidence has been presented for interactions of MurB or MurA in the complex as of yet, and we rarely found either gene in the cluster. This organization is near identical to that of *E. coli*, suggesting that this is the ancestral gene cluster of these two groups. The Bacilli and Lactobacilli only had a few genes of the cluster in this organization and the fast evolving Erysipeltrichia did not have any usual organization.

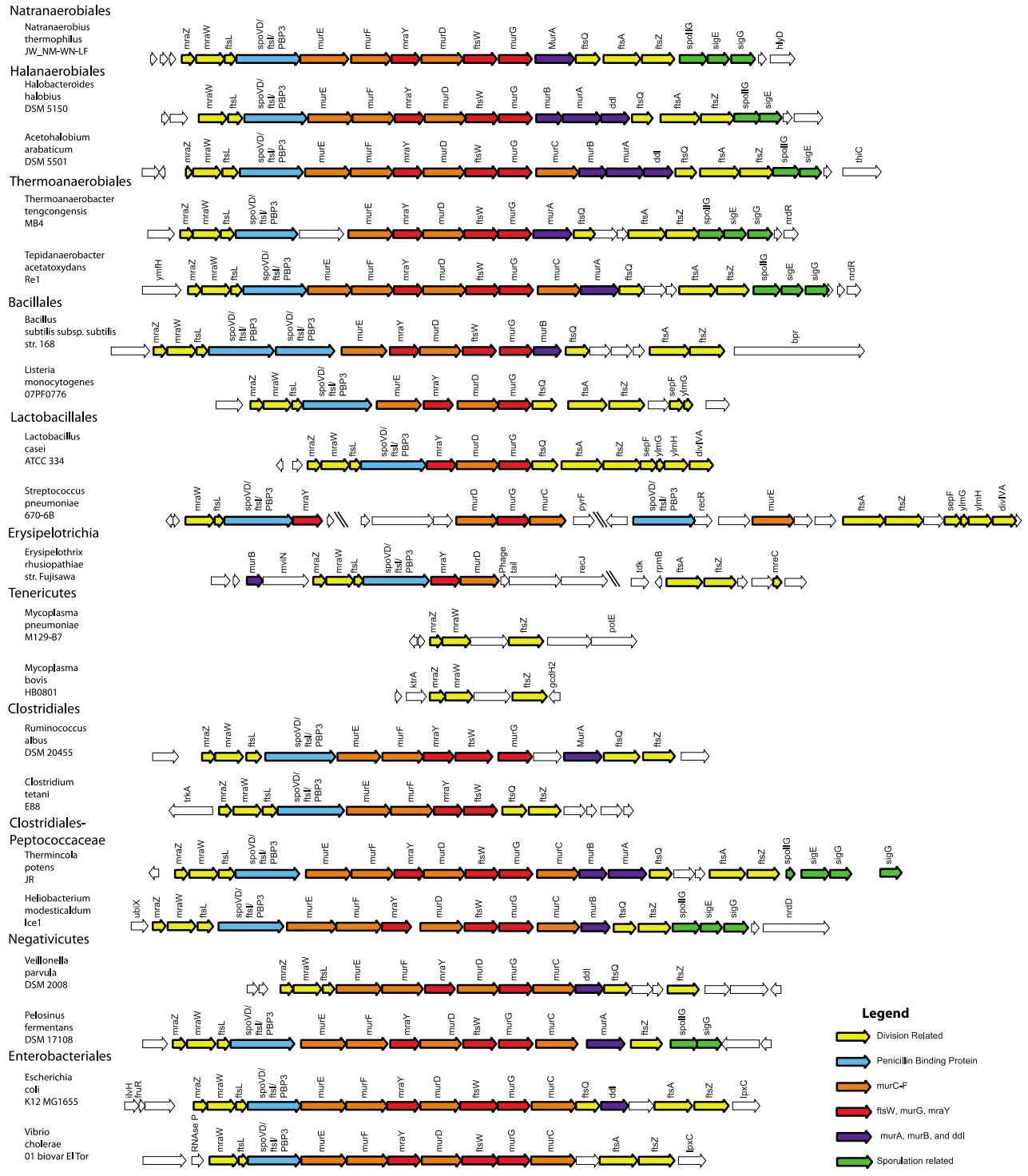

Fig 7. Local genomic area of PG biosynthetic gene cluster. Colours are noted in the legend. Double backslashes indicate large genomic distance.

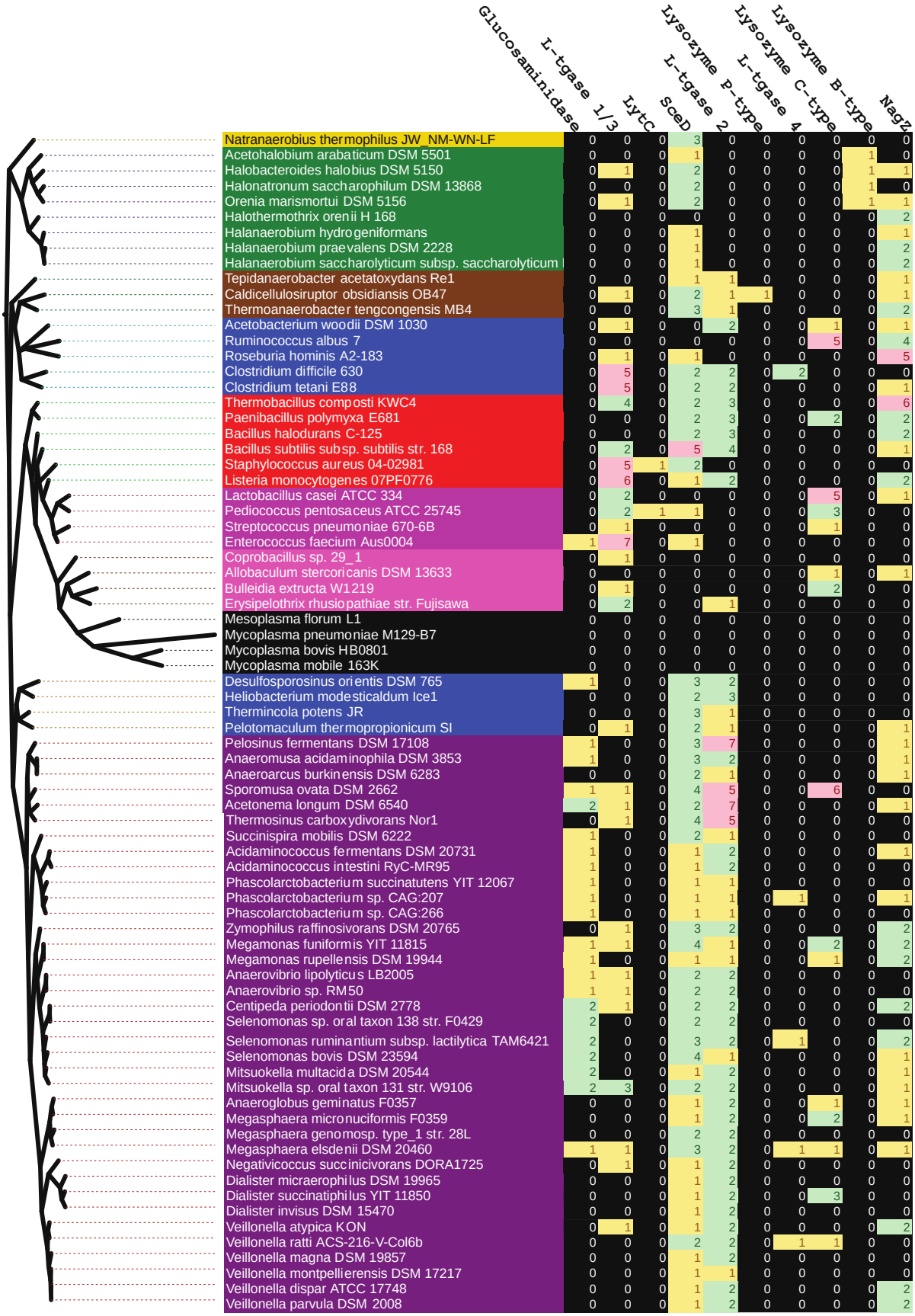

Fig 8. The copy number of PG backbone cleavage enzymes: Figure shows the total number of homologs found for each enzyme by HMM search. Colours behind numbers indicate copy number, 0=black, 1=yellow, 2-4=green >4=Red. Background colours correspond to individual orders.

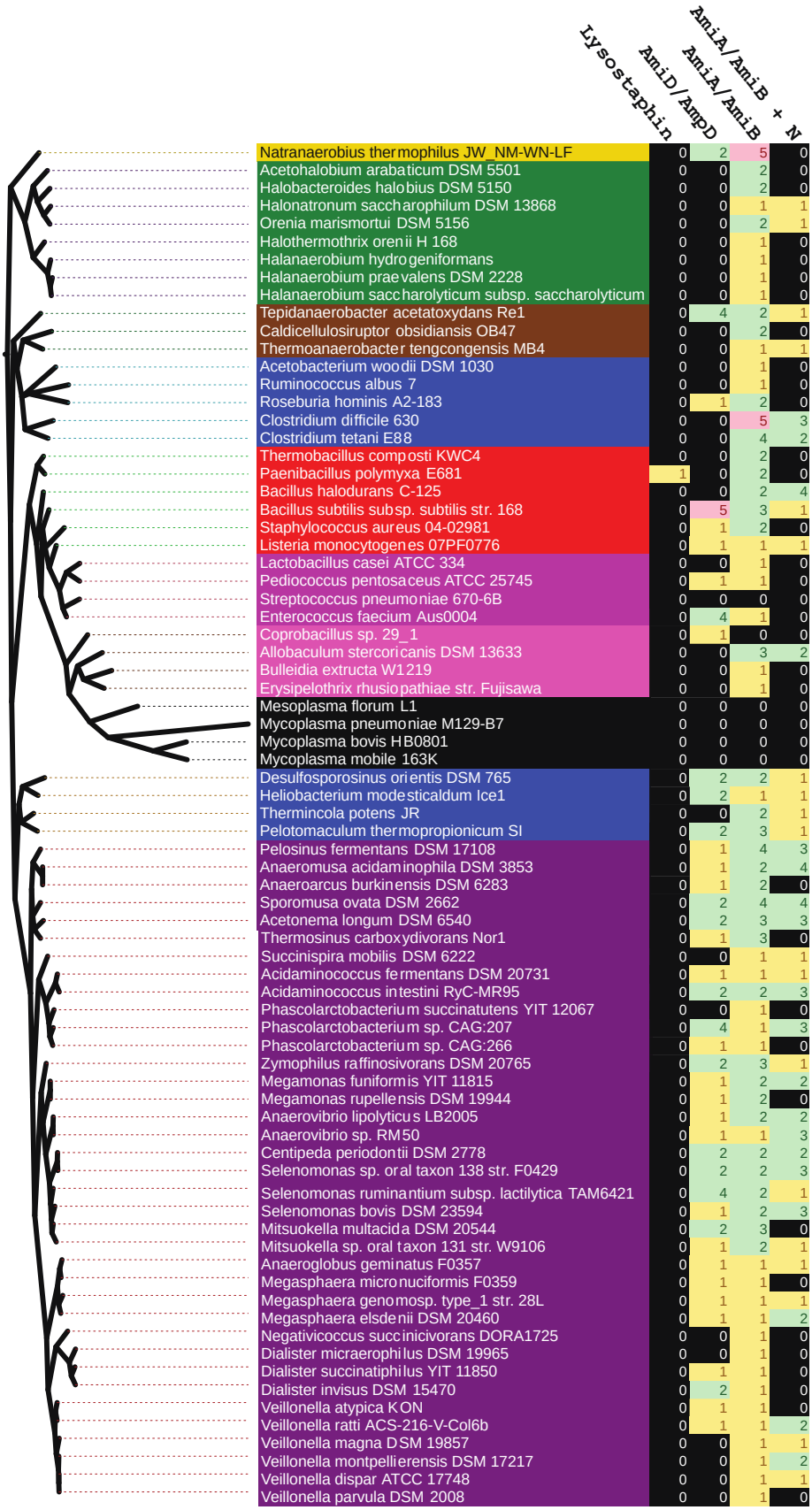

Fig 9. The copy number of PG amidases: Figure shows the total number of homologs found for each enzyme by HMM search. Colours behind numbers indicate copy number, 0=black, 1=yellow, 2-4=green >4=Red. Background colours correspond to individual orders.

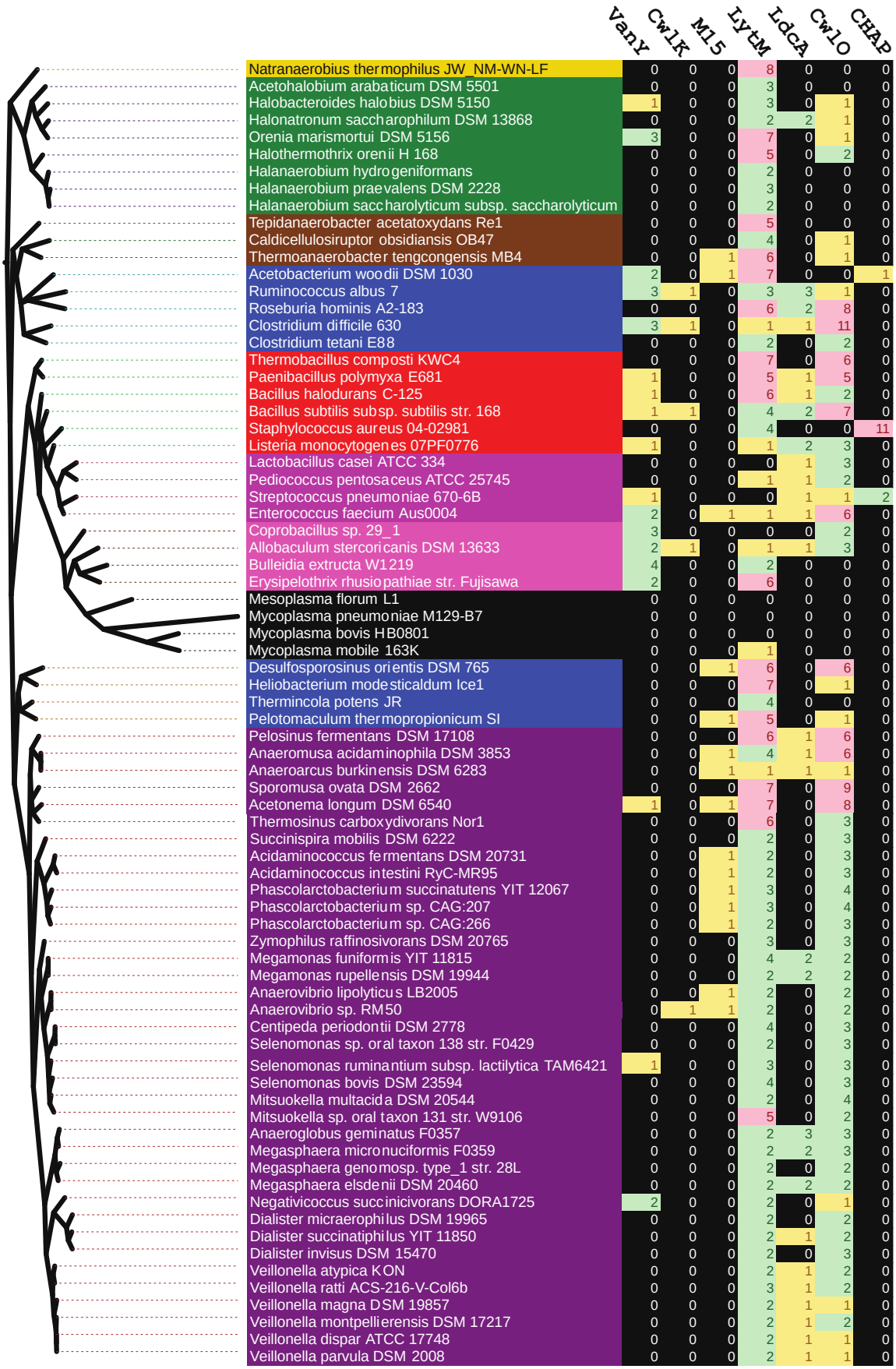

Fig 10. The copy number of PG peptidases: Figure shows the total number of homologs found for each enzyme by HMM search. Colours behind numbers indicate copy number, 0=black, 1=yellow, 2-4=green >4=Red. Background colours correspond to individual orders.

Lysozyme type enzymes. We continued our search for the enzyme(s) responsible for the difference in PG thickness by looking at enzymes responsible for cleaving the PG backbone (Fig 8). The true lysozymes, P, B, and C, were found in low copy number in only a few species. This is congruent with other bacteria as these enzymes are rarely used for cell PG maintenance (Humann and Lenz, 2009). Type 1 and 3 Lytic transglycosylases were found mostly in Bacilli and Clostridia, while the type 2 system was found predominantly in the Thermoanaerobacter, Negativicutes, but is also in some Clostridia and Bacilli. Type 4 lytic transglycosylases are usually prophage-encoded and we only found this type in a few genomes. It is interesting that the Halanaerobiales and Natranaerobiales are lacking any traditional class of lytic transglycosylase. SceD is a recently discovered enzyme which may be a lytic transglycosylase of a new type, however no enzymatic evidence has been done (Stapleton et al., 2007). This enzyme was found throughout most of the Firmicutes and may be the primary lytic transglycosylase of this clade.

Amidases. As we could not find any clear candidates from the lysozymes and lytic transglycosylases, we continued to look for our candidate in the amidases (Fig 9). Amidases cleave the PG peptide from the sugar backbone and all have the same enzymatic function; the differences among these enzymes is a different ancestry and fold. We searched for four classes of amidases: lysostaphin, LytC AmiD/AmpD, and AmiA/AmiB. AmiC would have been detected in the AmiA/ AmiB results, but its catalytic domain is not highly conserved and may not be detected. Lysostaphin was only found in one species and is a marker for antibiotic resistance. AmiD/AmpD was found in several species, but no clear pattern of monoderm/diderm absence or presence. AmiA/AmiB was found in high numbers in most of the bacteria. There was a clear divide in both phylogeny and in the alignment and this class was divided as to whether it contained an N-terminal domain or not. There were no detectable features of this additional domain and it did not clearly enlighten our solution. The CHAP catalytic domain found in LytC is usually part of a phage system.

Peptidases. We continued our search for the candidate enzyme by looking for peptidases (Figure 10). Peptidases are widespread throughout all life and as such we searched for those with known PG involvement. We searched for several domains and found that most were found sparingly in a few genomes. Many of these were previously shown to be phage-encoded (Layec et al., 2008). The two primary conserved peptidases, CwlO and LytM, are found in most, but not all, Firmicutes studied. There is no clear indication of involvement in the control of PG thickness.

Sporulation. We concluded our study with an analysis of the enzymes involved sporulation (Figure 11). We started our search by conducting a literature search and a search for the master regulator *spoOA* to determine which bacteria are spore formers. We found that several bacteria are likely to be spore formers, such as *N thermophilus,* as they have many genes, even though they are noted as non-sporulating in the literature. The δ-lactam forming enzyme CwlD was found in all spore formers and some non-spore formers, however, we could not search for PdaA as it was indistinguishable from its homologs without functional similarity. These and the opposing the degrading enzyme SleB was found in all spore formers and only one non-sporulating bacteria.
SpoIID and SpoIIP proved to be an anomaly. These two genes work in a 1:1 ratio to degrade the PG and move the second membrane around the daughter cell. Both enzymes were found in all sporulating bacteria with their corresponding helper enzymes SpoIIM and SpoIIIAH. The interesting fact is that *spoIID* is also found in non-sporulating Negativicutes without any other sporulating genes. The same is true of the non-sporulating Halanaerobiales, except they still have *spoOA* and *sleB*. This proved to be interesting so we went on to do a phylogeny of *spoIID* (Figure 12).
We took a two-pronged method to determine the function of SpoIID in the diderm Firmicutes; we did a phylogeny and looked at genomic syntany. We found that spoIID falls into one of three families: a sporulating class found only in sporulating bacteria, a divergent class, found to be phage-encoded or in several regions of different genomic syntany, and the diderm type. The diderm type was found in a few monoderm clostridia and all of the diderm Firmicutes.

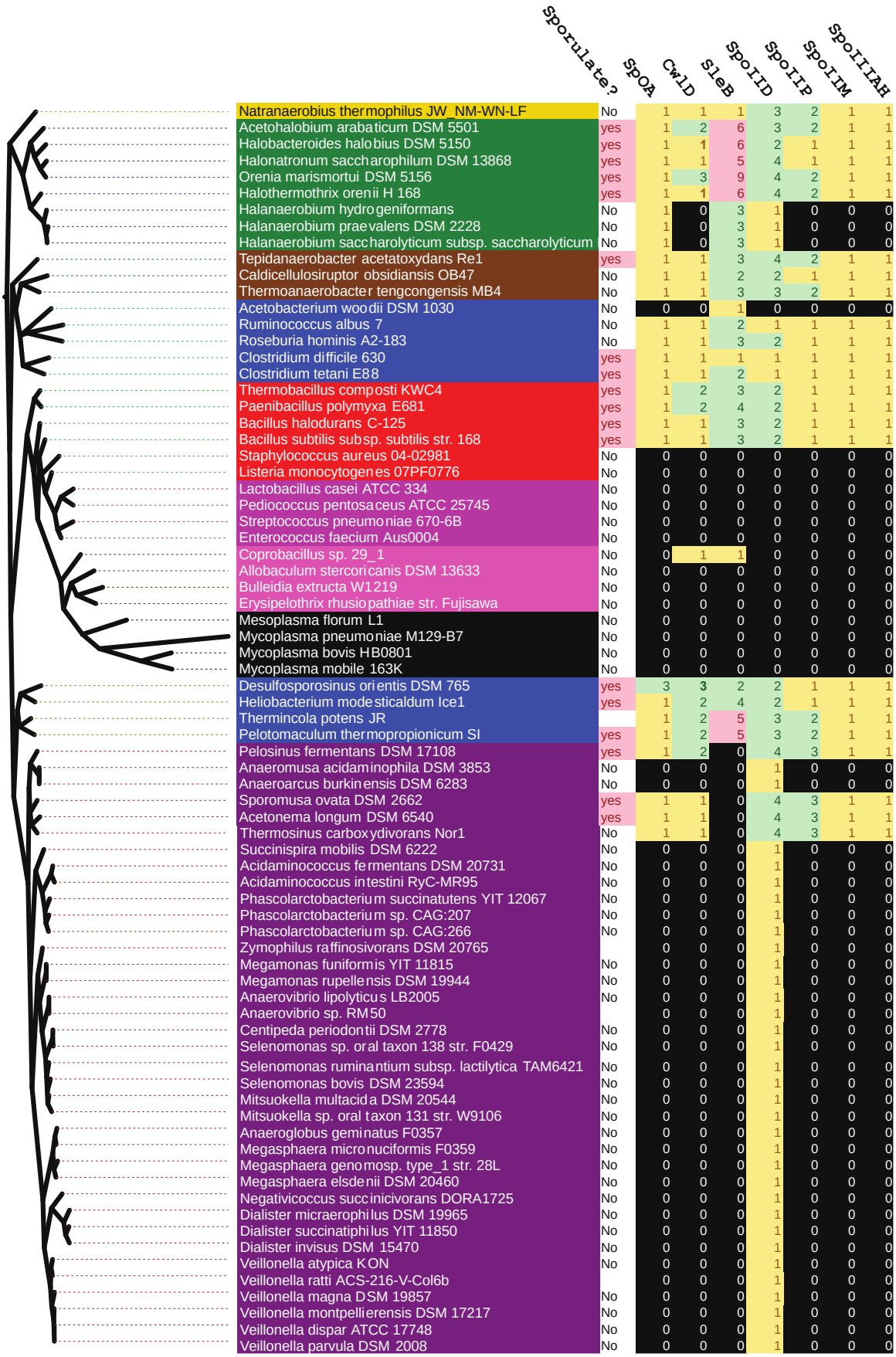

Fig 11. The copy number of Enzymes involved in PG remodelling in sporulation: Figure shows the total number of homologs found for each enzyme by HMM search. Colours behind numbers indicate copy number, 0=black, 1=yellow, 2-4=green >4=Red. Background colours correspond to individual orders.

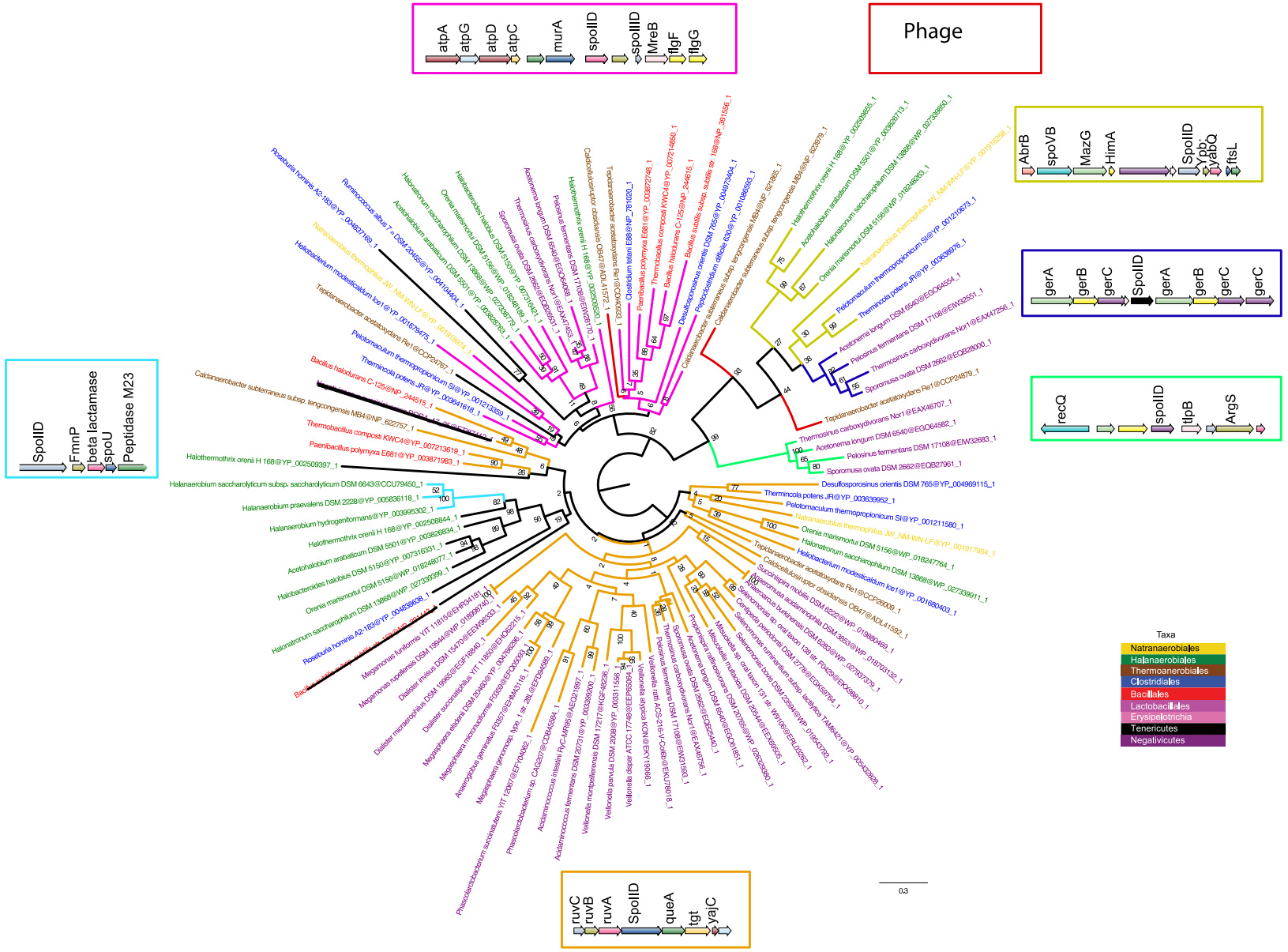

Fig 12 Phylogeny of Firmicute SpoIID The tree was constructed with PhyML using 160 residues and an LG model with a Gamma distribution. The colour of the outline of the syntany boxes corresponds to the colour of the branches it represents.

Discussion and Conclusions

The goals of this study were to determine if the PG of diderm Firmicutes had unique characteristics and to discover any catalytic remodelling domains that are found exclusively in the monoderm phyla or exclusively in the diderm phyla. We managed to successfully to meet both goals.

Initially, we found that the PG of the Firmicutes matched up to known stereotypes with the monoderm bacteria possessing thick PG and the diderm bacteria having much thinner PG, similar to *E. coli* (Matias and Beveridge, 2006). These data were unsurprising but necessary as no similar comparison could be found in the literature.

We then proceeded to look at the copy number of the core biosynthetic machinery and construct phylogenies when an odd pattern occurs. Our results matched well with the known components of the biosynthetic machinery of Bacteria (Sauvage *et al.*, 2008b; Gautam *et al.*, 2011), however, we did find a few novel results. Most Negativicutes only possessed a single PBPA protein but had an additional MGT protein. This MGT can only potentially perform the transglycosylation reaction and may replace the secondary copies of PBPA that are found in most bacteria (Sauvage *et al.*, 2008a). The other curious feature of the diderm Firmicutes was the presence of an additional homolog of FtsW. Three different paralogs of FtsW exist, one for sporulation, one for rod shape, and a final copy as the primary PG flippase (Mohammadi *et al.*, 2011). We could easily group the different types of the enzyme on *B. subtilis* homologs, however, there is no known function for the Negativicute FtsW homolog and may have a novel function.

In the end, the primary focus of this analysis was to identify a candidate, which may be responsible for the diderm/monoderm transition. Our analysis pointed to a single candidate spoIID. The ideal presence/absence in our scenario was a gene found in all diderm Firmicutes and absent from all monoderm Firmicutes. SpoIID did not quite meet that criteria, however, the copies that were found in the monoderms were all sporulating associated, while the non-sporulating diderm Firmicutes had a copy with no clear use in sporulation. Those diderm Firmicutes that do sporulate, such as *A. longum*, possess an additional copy beyond that of *spoIIP* the functional partner of SpoIID. This could mean that all diderm Firmicutes have this homolog for a non-canon function, that of maintaining a thin PG. The peptidoglycan regulation of *E. coli* is done through two OM lipoproteins, LpoA and LpoB, which coordinate the activity of the penicillin-binding proteins and regulate growth (Typas *et al.*, 2010). These proteins are not found outside the Proteobacteria (Data not shown), and another system must perform its function.

SpoIID immediately suggests a mechanism for PG thickness regulation as it is responsible for thinning and separating the PG and dragging an extra membrane layer during the incomplete division of sporulation (Gutierrez *et al.*, 2010). Known data on spoIID is covered in depth in the next chapter of this thesis.

Materials and Methods

Genomic local databank. We assembled a local databank of seventy-six genomes representing each of the seven orders available, with an enrichment of the Negativicutes and Halanaerobiales.

Table 1. PFAM models used in PG enzyme search: Known PFAM domains corresponding to PG remodelling activity

PFAM ID	Name of PFAM domain
PF06725	3D
PF01510	Amidase_2
PF01520	Amidase_3
PF05382	Amidase_5
PF12123	Amidase02_C
PF00144	Beta-lactamase
PF05257	CHAP
PF07478	Dala_Dala_lig_C
PF01820	Dala_Dala_lig_N
PF06335	DUF1054
PF02388	FemAB
PF01832	Glucosaminidase
PF05838	Glyco_hydro_108
PF01183	Glyco_hydro_25
PF00933	Glyco_hydro_3
PF04965	GPW_gp25
PF07486	Hydrolase_2
PF00753	Lactamase_B
PF00062	Lys
PF00877	NLPC_P60
PF00246	Peptidase_M14
PF01427	Peptidase_M15
PF13539	Peptidase_M15_4
PF01551	Peptidase_M23
PF03411	Peptidase_M74
PF00768	Peptidase_S11
PF02113	Peptidase_S13
PF02016	Peptidase_S66
PF01464	SLT
PF13406	SLT_2
PF08486	SpoIID
PF07454	SpoIIP
PF00912	Transgly
PF06737	Transglycosylas
PF00905	Transpeptidase
PF14814	UB2H
PF02557	VanY
PF03734	YkuD
PF00959	Phage_lysozyme

HMM construction and use. Core PG biosynthetic proteins (MurA-G, MurJ, MraY, FtsW) were searched for using a manually created HMM (Hidden Markov Model). Genes for HMM construction were obtained using *Bacillus subtilis* sub species *subtilis* 168 as a query. These genes were queried in the non-redundant database at NCBI with BLAST(Altschul *et al.*, 1990). The top twenty genes from distinct genera were chosen and an HMM profile was constructed using HMMER2 with default settings(Mistry *et al.*, 2013).

Searches for active PG remodelling enzymes was performed using all known PFAM HMM models (Finn *et al.*, 2014). These models were obtained by three methods. First, all PFAMs were taken from two other studies(Layec *et al.*, 2008; Machowski *et al.*, 2014). A

second method was to query the PFAM website with all known PG modifying enzymes (Appendix A). The final method was to do a text query of the PFAM database for databases related to the words "peptidoglycan", "murien", "penicillin", "lysozyme", "lytic transglycosylase", and "cell wall". This dataset was reduced to only the domains with known catalytic function. A full list of active site domains is listed in Table 1.1. Some search results were further characterized by the presence/absence of a second domain. These criteria can be found in Table 1.2.
Ribosomal proteins were searched using PFAMs for known domains found in each of the individual ribosomal proteins. Out of an initial set of 53 ribosomal proteins, 6 were removed due to non-universal distribution (L25, L30, S4, S14, S21) or having various paralogs (L33, S4, S14).
Searches were conducted using HMMER2 on default settings. The selection was further refined after alignment with MUSCLE v3.8.31 (Edgar, 2004). Sequences that did not align removed.

Tree Construction. Alignments selected for phylogenetic reconstruction were trimmed using BMGE 1.1 (Criscuolo and Gribaldo, 2010b) with a BLOSSUM30 substitution matrix.
Maximum likelihood (ML) trees were calculated by PhyML 3.1(Guindon et al., 2010), with the PROTGAMMAILG model and four categories of evolutionary rates, as suggested by the Akaike Information Criterion (AIC) implemented in ProtTest 3 (Darriba et al., 2011), and by PhyML 3.1, with the same model. Branch robustness was estimated with the nonparametric bootstrap procedure implemented in PhyML (100 replicates of the original dataset).
A separate ML analysis was conducted using RaxML 8 (Stamatakis, 2014) using the same parameters as PhyML for independent validation of our datasets.
PhyloBayes 3. 3(Lartillot et al., 2009) was used to perform Bayesian analysis using the CAT+GTR models and a gamma distribution with four categories of evolutionary rates was used to model the heterogeneity of site evolutionary rates. For each dataset, two independent chains were run until convergence as stated in the documentation. The first 25% of trees were discarded as burn-in and the posterior consensus was computed by selecting one out of every two trees to compute the 50% majority consensus tree.
Determination of peptidoglycan thickness
A complete literature review was conducted to determine PG thickness in all species studied. Many species lacked data, and only a few had accurate measurements. When only Negative staining EM pictures were available, the thickness was measured using the scale bars or by the magnification.

Table 2. Criteria for protein classification

Protein class	Known Proteins	Primary search term	Secondary Search term
Core	MurA	Constructed HMM	
	MurB	Constructed HMM	
	MurC	Constructed HMM	
	MurD	Constructed HMM	
	MurE	Constructed HMM	
	MurF	Constructed HMM	
	Ddl	Dala_Dala_lig_C (PF07478)	Dala_Dala_lig_N (PF01820)
	MurG	Constructed HMM	
	MraY	Constructed HMM	
	MurJ	Constructed HMM	
PBP Class A	PBP1a, PBP1b, PBP1c	Transgly (PF00912)	Transpeptidase (PF00905)
MGT	MGTs	Transgly (PF00912)	No other domain
PBP Class B	PBP2, PBP3	Transpeptidase (PF00905)	PDB_dimer (PF03717)
PBP Class C AMP-H	AMP-H	Beta-Lactamase (PF00144)	Excluding other Class C PBP
PBP Class C4	DacC	Peptidase_S13(PF02113)	
PBP Class C5	DacA, DacB, DacF	Peptidase_S11 (PF00768)	PBP5_C (PF07943)
PBP Class C7	PBP7	Peptidase_S11 (PF00768)	Without PBP5_C (PF07943)
MurNAc-l-Ala amidases	Lysostaphin	Amidase_5 (PF05382)	
	AmiD,AmpD	Amidase_2 (PF01510)	
	AmiA/AmiB	Amidase_3 (PF01520)	
	AmiA/AmiB +N	Amidase_3 (PF01520)	N-termini domain
	LytC	CHAP (PF05257)	
L-D transpeptidase	ErfK/YbiS/YcfS/YnhG,YkuD	YkuD (PF03734)	
Crosslink cleavage (D-ala, D-ala)	VanY	VanY (PF02557)	
	CwlK	M-15-4 (PF13539)	
	M15	M15 (PF01427)	
L,D Carboxypeptidase	ldcA	Peptidase_S66 (PF02016)	
Endopeptidase	LytM	Peptidase_M23 (PF01551)	
	CwlO,CwlS,LytF,lytE	NLPC_P60 (PF00877)	
N-acetyl glucosamididase	LytC	glucosamididase (PF01832)	SH3 (CL0010)
	Glucosamididase	glucosamididase (PF01832)	Without SH3
Lytic Transglycosylase	Family 1 /3	SLT (PF01464)	
	Family 2	3D (PF06725)	
	Family 4	Phage_lysozyme (PF00959)	
Lysozyme	Fungal Type (B)	GPW_gp25(PF04965)	
	Chicken Type (C)	Glyco_hydro_25 (PF01183)	
	Goose Type (P)	Lys (PF00062)	
Bacterial Type	NagZ	Glyco_hydro_3 (PF00933)	
	SceD	Glyco_hydro_108 (PF05838)	
Secreted Bacterial	SpoIIIAH	SpoIIIAH	
	SpoIIM	SpoIIM	
	SpoIIP	SpoIIP (PF07454)	
	SpoIID	SpoIID (PF08486)	
	SleB/CwlJ	Hydrolase_2 (PF07486)	
	FlgJ		

Table 3. References for PG thickness data

Species	Source
Acetonema longum DSM 6540	(Tocheva *et al.*, 2011)
Acidaminococcus fermentans DSM 20731	(Rogosa, 1969)
Acidaminococcus intestini RyC-MR95	(Jumas-Bilak *et al.*, 2007)
Anaeroglobus geminatus F0357	(Carlier *et al.*, 2002)
Anaerovibrio lipolyticus LB2005	(Henderson and Hodgkiss, 1973)
Centipeda periodontii DSM 2778	(Males *et al.*, 1984)
Dialister invisus DSM 15470	(Downes *et al.*, 2003)
Dialister micraerophilus DSM 19965	(Jumas-Bilak *et al.*, 2005)
Dialister succinatiphilus YIT 11850	(Morotomi *et al.*, 2008)
Megamonas funiformis YIT 11815	(Chevrot *et al.*, 2008)
Megasphaera elsdenii DSM 20460	(Costerton *et al.*, 1974)
Megasphaera micronuciformis F0359	(Marchandin *et al.*, 2003)
Mitsuokella multacida DSM 20544	(Kalmokoff *et al.*, 2009)
Negativicoccus succinicivorans DORA1725	(Marchandin *et al.*, 2010)
Selenomonas bovis DSM 23594	(Zhang and Dong, 2009)
Selenomonas ruminantium subsp. lactilytica TAM6421	(Fulghum and Moore, 1963)
Sporomusa ovata DSM 2662	(*Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*, 2011)
Thermosinus carboxydivorans Nor1	(Lee *et al.*, 2006)
Veillonella dispar ATCC 17748	(*Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*, 2011)
Veillonella montpellierensis DSM 17217	
Veillonella parvula DSM 2008	(Unpublished data)
Acetohalobium arabaticum DSM 5501	(*Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*, 2011)
Halanaerobium hydrogeniformans	(*Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*, 2011)
Halanaerobium praevalens DSM 2228	(Zeikus, Hegge, Thompson, Phelps, and T. a. Langworthy, 1983)
Halanaerobium saccharolyticum subsp. saccharolyticum DSM 6643	(Zeikus, Hegge, Thompson, Phelps, and T. A. Langworthy, 1983)
Halothermothrix orenii H 168	(Cayol *et al.*, 1994)
Bacillus subtilis subsp. subtilis str. 168	(Matias and Beveridge, 2008)
Listeria monocytogenes 07PF0776	(Goldfine and Shen, 2007)
Paenibacillus polymyxa E681	(*Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*, 2011)
Staphylococcus aureus 04-02981	(Matias and Beveridge, 2006)
Caldicellulosiruptor obsidiansis OB47	(Hamilton-Brehm *et al.*, 2010)
Thermoanaerobacter tengcongensis MB4	(Xue *et al.*, 2001)
Pelotomaculum thermopropionicum SI	(Imachi *et al.*, 2002)
Thermincola potens JR	(*Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*, 2011)
Enterococcus faecium Aus0004	(Higgins and Shockman, 1970)
Lactobacillus casei ATCC 334	(Regulski *et al.*, 2012)
Streptococcus pneumoniae 670-6B	(Hammerschmidt *et al.*, 2005)
Acetobacterium woodii DSM 1030	(Mayer *et al.*, 1977)
Clostridium difficile 630	(Permpoonpattana *et al.*, 2013)
Clostridium tetani E88	(Takagi *et al.*, 1965)
Ruminococcus albus 7	(Morrison and Miron, 2000)
Natranaerobius thermophilus JW_NM-WN-LF	(Mesbah *et al.*, 2007)
Bulleidia extructa W1219	(Downes *et al.*, 2000)
Erysipelothrix rhusiopathiae str. Fujisawa	(Shi *et al.*, 2012)
Mesoplasma florum L1	(*Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*, 2011)
Mycoplasma bovis HB0801	(*Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*, 2011)
Mycoplasma mobile 163K	(*Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*, 2011)
Mycoplasma pneumoniae M129-B7	(*Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*, 2011)

Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes. (2011). Springer; 2 edition.
Al-Hinai, M. A., Jones, S. W., & Papoutsakis, E. T. (2015). The Clostridium sporulation programs: diversity and preservation of endospore differentiation. Microbiology and Molecular Biology Reviews : MMBR, 79(1), 19–37. http://doi.org/10.1128/MMBR.00025-14
Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–10. http://doi.org/10.1016/S0022-2836(05)80360-2
Barreteau, H., Kovac, A., Boniface, A., Sova, M., Gobec, S., & Blanot, D. (2008). Cytoplasmic steps of peptidoglycan biosynthesis. FEMS Microbiology Reviews, 32(2), 168–207. http://doi.org/10.1111/j.1574-6976.2008.00104.x
Bouhss, A., Trunkfield, A. E., Bugg, T. D. H., & Mengin-Lecreulx, D. (2008). The biosynthesis of peptidoglycan lipid-linked intermediates. FEMS Microbiology Reviews, 32(2), 208–33. http://doi.org/10.1111/j.1574-6976.2007.00089.x
Carlier, J.-P., Marchandin, H., Jumas-Bilak, E., Lorin, V., Henry, C., Carrière, C., & Jean-Pierre, H. (2002). Anaeroglobus geminatus gen. nov., sp. nov., a novel member of the family Veillonellaceae. International Journal of Systematic and Evolutionary Microbiology, 52(Pt 3), 983–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12054267
Cava, F., & de Pedro, M. A. (2014). Peptidoglycan plasticity in bacteria: emerging variability of the murein sacculus and their associated biological functions. Current Opinion in Microbiology, 18, 46–53. http://doi.org/10.1016/j.mib.2014.01.004
Cayol, J. L., Ollivier, B., Patel, B. K. C., Prensier, G., Guezennec, J., & Garcia, J. L. (1994). Isolation and Characterization of Halothemothrix orenii gen . International Journal of Systematic Bacteriology, 44(3), 534–540. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7520742
Chevrot, R., Carlotti, A., Sopena, V., Marchand, P., & Rosenfeld, E. (2008). Megamonas rupellensis sp. nov., an anaerobe isolated from the caecum of a duck. International Journal of Systematic and Evolutionary Microbiology, 58(12), 2921–2924. http://doi.org/10.1099/ijs.0.2008/001297-0
Costerton, J. W., Damgaard, H. N., & Cheng, K. J. (1974). Cell envelope morphology of rumen bacteria. Journal of Bacteriology, 118(3), 1132–43. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=246864&tool=pmcentrez&rendertype=abstract
Criscuolo, A., & Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evolutionary Biology, 10, 210. http://doi.org/10.1186/1471-2148-10-210
Darriba, D., Taboada, G. L., & Posada, D. (2011). Supplementary material for " ProtTest 3 : fast selection of best-fit models of protein evolution " Summary : Availability :, 1–4. http://doi.org/10.1093/bioinformatics/btr088
Downes, J., Munson, M., & Wade, W. G. (2003). Dialister invisus sp. nov., isolated from the human oral cavity. International Journal of Systematic and Evolutionary Microbiology, 53(Pt 6), 1937–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14657126
Downes, J., Olsvik, B., Hiom, S. J., Spratt, D. A., Cheeseman, S. L., Olsen, I., … Wade, W. G. (2000). Bulleidia extructa gen. nov., sp. nov., isolated from the oral cavity. International Journal of Systematic and Evolutionary Microbiology, 50 Pt 3, 979–83. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10843035
Dubochet, J., McDowall, A. W., Menge, B., Schmid, E. N., & Lickfeld, K. G. (1983). Electron microscopy of frozen-hydrated bacteria. J. Bacteriol., 155(1), 381–390. Retrieved from http://jb.asm.org/content/155/1/381.abstract?ijkey=6cb54f7744196d1e7831c69254b47ea078d0cce7&keytype2=tf_ipsecsha
Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792–1797. http://doi.org/10.1093/nar/gkh340
Eraso, J. M., Markillie, L. M., Mitchell, H. D., Taylor, R. C., Orr, G., & Margolin, W. (2014). The highly conserved MraZ protein is a transcriptional regulator in Escherichia coli. Journal of Bacteriology, 196(11), 2053–66. http://doi.org/10.1128/JB.01370-13
Ercoli, G., Tani, C., Pezzicoli, A., Vacca, I., Martinelli, M., Pecetta, S., … Aricò, B. (2015). LytM proteins play a crucial role in cell separation, outer membrane composition, and pathogenesis in nontypeable Haemophilus influenzae. mBio, 6(2), e02575. http://doi.org/10.1128/mBio.02575-14
Favini-Stabile, S., Contreras-Martel, C., Thielens, N., & Dessen, A. (2013). MreB and MurG as scaffolds for the cytoplasmic steps of peptidoglycan biosynthesis. Environmental Microbiology, 15(12), 3218–3228. http://doi.org/10.1111/1462-2920.12171
Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., … Punta, M. (2014). Pfam: the protein families database. Nucleic Acids Research, 42(Database issue), D222–30. http://doi.org/10.1093/nar/gkt1223
FULGHUM, R. S., & MOORE, W. E. (1963). ISOLATION, ENUMERATION, AND CHARACTERISTICS OF PROTEOLYTIC RUMINAL BACTERIA. Journal of Bacteriology, 85, 808–15. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=278229&tool=pmcentrez&rendertype=abstract
Gautam, A., Vyas, R., & Tewari, R. (2011). Peptidoglycan biosynthesis machinery: a rich source of drug targets. Critical Reviews in Biotechnology, 31(4), 295–336. http://doi.org/10.3109/07388551.2010.525498
Gilmore, M. E., Bandyopadhyay, D., Dean, A. M., Linnstaedt, S. D., & Popham, D. L. (2004). Production of muramic delta-lactam in Bacillus subtilis spore peptidoglycan. Journal of Bacteriology, 186(1), 80–9. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcg

i?artid=303458&tool=pmcentrez&rendertype=abstract
Goldfine, H., & Shen, H. (Eds.). (2007). Listeria monocytogenes: Pathogenesis and Host Response. Boston, MA: Springer US. http://doi.org/10.1007/978-0-387-49376-3
Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology, 59(3), 307–21. http://doi.org/10.1093/sysbio/syq010
Gutierrez, J., Smith, R., & Pogliano, K. (2010). SpoIID-mediated peptidoglycan degradation is required throughout engulfment during Bacillus subtilis sporulation. Journal of Bacteriology, 192(12), 3174–86. http://doi.org/10.1128/JB.00127-10
Hamilton-Brehm, S. D., Mosher, J. J., Vishnivetskaya, T., Podar, M., Carroll, S., Allman, S., … Elkins, J. G. (2010). Caldicellulosiruptor obsidiansis sp. nov., an anaerobic, extremely thermophilic, cellulolytic bacterium isolated from Obsidian Pool, Yellowstone National Park. Applied and Environmental Microbiology, 76(4), 1014–20. http://doi.org/10.1128/AEM.01903-09
Hammerschmidt, S., Wolff, S., Hocke, A., Rosseau, S., Müller, E., & Rohde, M. (2005). Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. Infection and Immunity, 73(8), 4653–67. http://doi.org/10.1128/IAI.73.8.4653-4667.2005
Hashimoto, M., Ooiwa, S., & Sekiguchi, J. (2012). Synthetic lethality of the lytE cwlO genotype in Bacillus subtilis is caused by lack of D,L-endopeptidase activity at the lateral cell wall. Journal of Bacteriology, 194(4), 796–803. http://doi.org/10.1128/JB.05569-11
Henderson, C., & Hodgkiss, W. (1973). An electron microscopic study of Anaerovibrio lipolytica (strain 5S) and its lipolytic enzyme. Journal of General Microbiology, 76(2), 389–93. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4124712
Higgins, M. L., & Shockman, G. D. (1970). Model for Cell Wall Growth of Streptococcus faecalis. J. Bacteriol., 101(2), 643–648. Retrieved from http://jb.asm.org/content/101/2/643?ijkey=bd37d7f9f270018e6db171f9371cec41fae2ba65&keytype2=tf_ipsecsha
Hilbert, D. W., & Piggot, P. J. (2004). Compartmentalization of gene expression during Bacillus subtilis spore formation. Microbiology and Molecular Biology Reviews : MMBR, 68(2), 234–62. http://doi.org/10.1128/MMBR.68.2.234-262.2004
Humann, J., & Lenz, L. L. (2009). Bacterial peptidoglycan degrading enzymes and their impact on host muropeptide detection. Journal of Innate Immunity, 1(2), 88–97. http://doi.org/10.1159/000181181
Imachi, H., Sekiguchi, Y., Kamagata, Y., Hanada, S., Ohashi, A., & Harada, H. (2002). Pelotomaculum thermopropionicum gen. nov., sp. nov., an anaerobic, thermophilic, syntrophic propionate-oxidizing bacterium. International Journal of Systematic and Evolutionary Microbiology, 52(Pt 5), 1729–35. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12361280
Jacquier, N., Viollier, P. H., & Greub, G. (2015). The role of peptidoglycan in chlamydial cell division: towards resolving the chlamydial anomaly. FEMS Microbiology Reviews, 39(2), 262–275. http://doi.org/10.1093/femsre/fuv001
Jumas-Bilak, E., Carlier, J.-P., Jean-Pierre, H., Mory, F., Teyssier, C., Gay, B., … Marchandin, H. (2007). Acidaminococcus intestini sp. nov., isolated from human clinical samples. International Journal of Systematic and Evolutionary Microbiology, 57(Pt 10), 2314–9. http://doi.org/10.1099/ijs.0.64883-0
Jumas-Bilak, E., Jean-Pierre, H., Carlier, J.-P., Teyssier, C., Bernard, K., Gay, B., … Marchandin, H. (2005). Dialister micraerophilus sp. nov. and Dialister propionicifaciens sp. nov., isolated from human clinical samples. International Journal of Systematic and Evolutionary Microbiology, 55(Pt 6), 2471–8. http://doi.org/10.1099/ijs.0.63715-0
Kalmokoff, M. L., Austin, J. W., Cyr, T. D., Hefford, M. A., Teather, R. M., & Selinger, L. B. (2009). Physical and genetic characterization of an outer-membrane protein (OmpM1) containing an N-terminal S-layer-like homology domain from the phylogenetically Gram-positive gut anaerobe Mitsuokella multacida. Anaerobe, 15(3), 74–81. http://doi.org/10.1016/j.anaerobe.2009.01.001
Karamanos, Y. (1997). Endo-N-acetyl-beta-D-glucosaminidases and their potential substrates: structure/function relationships. Research in Microbiology, 148(8), 661–71. http://doi.org/10.1016/S0923-2508(99)80065-5
Kojima, S., & Kamio, Y. (2012). Molecular basis for the maintenance of envelope integrity in Selenomonas ruminantium: cadaverine biosynthesis and covalent modification into the peptidoglycan play a major role. Journal of Nutritional Science and Vitaminology, 58(3), 153–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22878384
Lartillot, N., Lepage, T., & Blanquart, S. (2009). PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics, 25(17), 2286–2288. http://doi.org/10.1093/bioinformatics/btp368
Layec, S., Decaris, B., & Leblond-Bourget, N. (2008). Diversity of Firmicutes peptidoglycan hydrolases and specificities of those involved in daughter cell separation. Research in Microbiology, 159(7-8), 507–15. http://doi.org/10.1016/j.resmic.2008.06.008
Leaver, M., Domínguez-Cuevas, P., Coxhead, J. M., Daniel, R. A., & Errington, J. (2009). Life without a wall or division machine in Bacillus subtilis. Nature, 457(7231), 849–53. http://doi.org/10.1038/nature07742
Lee, Y. J., Romanek, C. S., Mills, G. L., Davis, R. C., Whitman, W. B., & Wiegel, J. (2006). Gracilibacter thermotolerans gen. nov., sp. nov., an anaerobic, thermotolerant bacterium from a constructed wetland receiving acid sulfate water. International Journal of Systematic and Evolutionary Microbiology, 56(9), 2089–2093. http://doi.org/10.1099/ijs.0.64040-0
Machowski, E. E., Senzani, S., Ealand, C., & Kana, B. D. (2014). Comparative genomics for mycobacterial peptidoglycan remodelling enzymes reveals extensive genetic multiplicity. BMC Microbiology, 14(1), 75. http://doi.org/10.1186/1471-2180-14-75

Males, B. M., Berthold, P., Dougherty, P. A., & Listgarten, M. A. (1984). Helical flagellation in Centipeda periodontii, a gram-negative, anaerobic bacillus from periodontitis lesions. Journal of General Microbiology, 130(1), 185–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6707608
Marchandin, H., Jumas-Bilak, E., Gay, B., Teyssier, C., Jean-Pierre, H., de Buochberg, M. S., … Carlier, J.-P. (2003). Phylogenetic analysis of some Sporomusa sub-branch members isolated from human clinical specimens: description of Megasphaera micronuciformis sp. nov. International Journal of Systematic and Evolutionary Microbiology, 53(Pt 2), 547–53. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12710625
Marchandin, H., Teyssier, C., Campos, J., Jean-Pierre, H., Roger, F., Gay, B., … Jumas-Bilak, E. (2010). Negativicoccus succinicivorans gen. nov., sp. nov., isolated from human clinical samples, emended description of the family Veillonellaceae and description of Negativicutes classis nov., Selenomonadales ord. nov. and Acidaminococcaceae fam. nov. in the ba. International Journal of Systematic and Evolutionary Microbiology, 60(6), 1271–1279. http://doi.org/10.1099/ijs.0.013102-0
Matias, V. R. F., & Beveridge, T. J. (2006). Native cell wall organization shown by cryo-electron microscopy confirms the existence of a periplasmic space in Staphylococcus aureus. Journal of Bacteriology, 188(3), 1011–21. http://doi.org/10.1128/JB.188.3.1011-1021.2006
Matias, V. R. F., & Beveridge, T. J. (2008). Lipoteichoic acid is a major component of the Bacillus subtilis periplasm. Journal of Bacteriology, 190(22), 7414–8. http://doi.org/10.1128/JB.00581-08
Mayer, F., Lurz, R., & Schoberth, S. (1977). Electron microscopic investigation of the hydrogen-oxidizing acetate-forming anaerobic bacterium Acetobacterium woodii. Archives of Microbiology, 115(2), 207–13. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/596994
Mesbah, N. M., Hedrick, D. B., Peacock, A. D., Rohde, M., & Wiegel, J. (2007). Natranaerobius thermophilus gen. nov., sp. nov., a halophilic, alkalithermophilic bacterium from soda lakes of the Wadi An Natrun, Egypt, and proposal of Natranaerobiaceae fam. nov. and Natranaerobiales ord. nov. International Journal of Systematic and Evolutionary Microbiology, 57(Pt 11), 2507–12. http://doi.org/10.1099/ijs.0.65068-0
Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., & Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Research, 41(12), e121. http://doi.org/10.1093/nar/gkt263
Mohammadi, T., Karczmarek, A., Crouvoisier, M., Bouhss, A., Mengin-Lecreulx, D., & den Blaauwen, T. (2007). The essential peptidoglycan glycosyltransferase MurG forms a complex with proteins involved in lateral envelope growth as well as with proteins involved in cell division in Escherichia coli. Molecular Microbiology, 65(4), 1106–21. http://doi.org/10.1111/j.1365-2958.2007.05851.x
Mohammadi, T., van Dam, V., Sijbrandi, R., Vernet, T., Zapun, A., Bouhss, A., … Breukink, E. (2011). Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. The EMBO Journal, 30(8), 1425–32. http://doi.org/10.1038/emboj.2011.61
Morotomi, M., Nagai, F., Sakon, H., & Tanaka, R. (2008). Dialister succinatiphilus sp. nov. and Barnesiella intestinihominis sp. nov., isolated from human faeces. International Journal of Systematic and Evolutionary Microbiology, 58(Pt 12), 2716–20. http://doi.org/10.1099/ijs.0.2008/000810-0
Morrison, M., & Miron, J. (2000). Adhesion to cellulose by Ruminococcus albus : a combination of cellulosomes and Pil-proteins? FEMS Microbiology Letters, 185(2), 109–115. http://doi.org/10.1111/j.1574-6968.2000.tb09047.x
Pei, J., & Grishin, N. V. (2005). COG3926 and COG5526: a tale of two new lysozyme-like protein families. Protein Science : A Publication of the Protein Society, 14(10), 2574–81. http://doi.org/10.1110/ps.051656805
Permpoonpattana, P., Phetcharaburanin, J., Mikelsone, A., Dembek, M., Tan, S., Brisson, M.-C., … Cutting, S. M. (2013). Functional characterization of Clostridium difficile spore coat proteins. Journal of Bacteriology, 195(7), 1492–503. http://doi.org/10.1128/JB.02104-12
Pinho, M. G., Kjos, M., & Veening, J.-W. (2013). How to get (a)round: mechanisms controlling growth and division of coccoid bacteria. Nature Reviews. Microbiology, 11(9), 601–14. http://doi.org/10.1038/nrmicro3088
Regulski, K., Courtin, P., Meyrand, M., Claes, I. J. J., Lebeer, S., Vanderleyden, J., … Chapot-Chartier, M.-P. (2012). Analysis of the peptidoglycan hydrolase complement of Lactobacillus casei and characterization of the major γ-D-glutamyl-L-lysyl-endopeptidase. PloS One, 7(2), e32301. http://doi.org/10.1371/journal.pone.0032301
Rogosa, M. (1969). Acidaminococcus gen. n., Acidaminococcus fermentans sp. n., anaerobic gram-negative diplococci using amino acids as the sole energy source for growth. Journal of Bacteriology, 98(2), 756–66. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=284882&tool=pmcentrez&rendertype=abstract
Sauvage, E., Kerff, F., Terrak, M., Ayala, J. A., & Charlier, P. (2008). The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEMS Microbiology Reviews, 32(2), 234–58. http://doi.org/10.1111/j.1574-6976.2008.00105.x
Scheurwater, E., Reid, C. W., & Clarke, A. J. (2008). Lytic transglycosylases: bacterial space-making autolysins. The International Journal of Biochemistry & Cell Biology, 40(4), 586–91. http://doi.org/10.1016/j.biocel.2007.03.018
Schleifer, K. H., & Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriological Reviews, 36(4), 407–77. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=408328&tool=pmcentrez&rendertype=abstract
Schoonmaker, M. K., Bishai, W. R., & Lamichhane, G. (2014). Nonclassical transpeptidases of Mycobacterium tuberculosis alter cell size, morphology, the cytosolic matrix, protein localization,

virulence, and resistance to β-lactams. Journal of Bacteriology, 196(7), 1394–402. http://doi.org/10.1128/JB.01396-13
Sham, L.-T., Butler, E. K., Lebar, M. D., Kahne, D., Bernhardt, T. G., & Ruiz, N. (2014). Bacterial cell wall. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. Science (New York, N.Y.), 345(6193), 220–2. http://doi.org/10.1126/science.1254522
Shi, F., Harada, T., Ogawa, Y., Ono, H., Ohnishi-Kameyama, M., Miyamoto, T., … Shimoji, Y. (2012). Capsular polysaccharide of Erysipelothrix rhusiopathiae, the causative agent of swine erysipelas, and its modification with phosphorylcholine. Infection and Immunity, 80(11), 3993–4003. http://doi.org/10.1128/IAI.00635-12
Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics (Oxford, England), 30(9), 1312–3. http://doi.org/10.1093/bioinformatics/btu033
Stapleton, M. R., Horsburgh, M. J., Hayhurst, E. J., Wright, L., Jonsson, I.-M., Tarkowski, A., … Foster, S. J. (2007). Characterization of IsaA and SceD, two putative lytic transglycosylases of Staphylococcus aureus. Journal of Bacteriology, 189(20), 7316–25. http://doi.org/10.1128/JB.00734-07
Takagi, A., Nakamura, K., & Ueda, M. (1965). Electron microscope studies of the intracytoplasmic membrane system in Clostridium tetani and Clostridium botulinum. Japanese Journal of Microbiology, 9(3), 131–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5328024
Tocheva, E. I., Matson, E. G., Morris, D. M., Moussavi, F., Leadbetter, J. R., & Jensen, G. J. (2011). Peptidoglycan remodeling and conversion of an inner membrane into an outer membrane during sporulation. Cell, 146(5), 799–812. http://doi.org/10.1016/j.cell.2011.07.029
Turner, R. D., Vollmer, W., & Foster, S. J. (2014). Different walls for rods and balls: the diversity of peptidoglycan. Molecular Microbiology, 91(5), 862–74. http://doi.org/10.1111/mmi.12513
Typas, A., Banzhaf, M., Gross, C. A., & Vollmer, W. (2012). From the regulation of peptidoglycan synthesis to bacterial growth and morphology. Nature Reviews. Microbiology, 10(2), 123–36. http://doi.org/10.1038/nrmicro2677
Vollmer, W., Blanot, D., & de Pedro, M. A. (2008). Peptidoglycan structure and architecture. FEMS Microbiology Reviews, 32(2), 149–67. http://doi.org/10.1111/j.1574-6976.2007.00094.x
Vollmer, W., Joris, B., Charlier, P., & Foster, S. (2008). Bacterial peptidoglycan (murein) hydrolases. FEMS Microbiology Reviews, 32(2), 259–286. http://doi.org/10.1111/j.1574-6976.2007.00099.x
Xue, Y., Xu, Y., Liu, Y., Ma, Y., & Zhou, P. (2001). Thermoanaerobacter tengcongensis sp. nov., a novel anaerobic, saccharolytic, thermophilic bacterium isolated from a hot spring in Tengcong, China. International Journal of Systematic and Evolutionary Microbiology, 51(Pt 4), 1335–41. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11491330
Zeikus, J. G., Hegge, P. W., Thompson, T. E., Phelps, T. J., & Langworthy, T. a. (1983a). Isolation and description of Haloanaerobium praevalens gen. nov. and sp. nov., an obligately anaerobic halophile common to Great Salt Lake sediments. Current Microbiology, 9(4), 225–233. http://doi.org/10.1007/BF01567586
Zeikus, J. G., Hegge, P. W., Thompson, T. E., Phelps, T. J., & Langworthy, T. A. (1983b). Isolation and description ofHaloanaerobium praevalens gen. nov. and sp. nov., an obligately anaerobic halophile common to Great Salt Lake sediments. Current Microbiology, 9(4), 225–233. http://doi.org/10.1007/BF01567586
Zhang, K., & Dong, X. (2009). Selenomonas bovis sp. nov., isolated from yak rumen contents. International Journal of Systematic and Evolutionary Microbiology, 59(Pt 8), 2080–3. http://doi.org/10.1099/ijs.0.007641-0

3.2 Attempted mutation of spoIID in Veillonella parvula

After identifying spoIID as a candidate for the mechanism of monoderm/diderm transition I immediately wanted to mutate/express this gene as we might have been able to produce the transition in a laboratory setting. We had two possibilities, we could constitutively express spoIID in *B. subtilis* while the Bacterium underwent sporulation or we could try and mutate the conserved *spoIID* within a nonsporulating diderm Firmicute. At this point, we had concluded that the ancestral Firmicute envelope was diderm so we favoured the mutation route. In addition, a means of mutating *V. parvula* had been recently developed (discussed in depth in the following introduction).

Unfortunately, my efforts to mutate the *V. parvula* were unsuccessful and I could not obtain any mutants to study. As such the next section will be the shortest section with limited results and discussion section.

Attempted mutation of *spoIID* in *Veillonella parvula*

One of the most essential components of the Bacterial cell envelope is the peptidoglycan. This molecule is responsible for shape, structure, among other uses. Diderm (Gram-negative) bacteria possess thin peptidoglycan, one to two molecules thick, while monoderms (Gram-positive) have peptidoglycan that is several molecules thick. We recently discovered a candidate enzyme, which may be responsible for the diderm/monoderm transition. In this study, we attempted to mutate *spoIID* with both insertional and in frame deletion mutagenesis. Both methods failed, however insight was still obtained.

Peptidoglycan | Mutation | Sporulation | Evolution

In the previous section, I presented data that suggests that the diderm/monoderm transition occurred due to the loss of *spoIID*. SpoIID is an important enzyme of sporulation involved in the second stage when the daughter cell separates from the mother cell (Nocadello et al., 2016). Its enzymatic activity has been described as a lytic transglycosylase which cleaves between the NAM and the NAG of the PG. This cleavage produces a terminal 1,6-anhydro-NAM that can not be reformed by the penicillin binding proteins. It works in concert with SpoIIP, which cleaves the PG peptide side chain (Gutierrez et al., 2010). Furthermore, this enzyme is embedded in the membrane with an alpha helix and drags this membrane to form an OM. This immediately suggests a mechanism of OM formation and regulation in the diderm Firmicutes.

Mutation of *spoIID* is now possible within the diderm Firmicutes as a genetic system has recently been developed in the Qi laboratory in Oklahoma. To develop this system, they isolated *Veillonella* strains from saliva samples and then extracted plasmids from these isolates (Liu et al., 2012). They then modified this plasmid to produce an *E. coli* to *Veillonella* shuttle plasmid, JBSJL2. This plasmid was tested for competence efficiency and one strain *V. atypica* OK5 was found to be suitable.

Building on this work the laboratory went on to construct a counter-selectable markerless system for *V. atypica* OK5 (Figure 1) (Zhou et al., 2015). To use this genetic system, first, a fragment upstream and downstream of the target deletion must be selected and cloned into pBST-Pmdh-pheS* within *E. coli*. The plasmid is then electroporated into *V. atypica* where it can not replicate. The cells are then exposed to the antibiotic and those that have taken up the plasmid will show resistance and survive. When the bacteria are exposed to a second marker, phenylalanine analogue p-chloro-phenylalanine, which is toxic when metabolized by the PheS* protein. As such cells that survive will have lost the plasmid. One of two events can happen, the cell can revert to the native state or the desired deletion can occur.

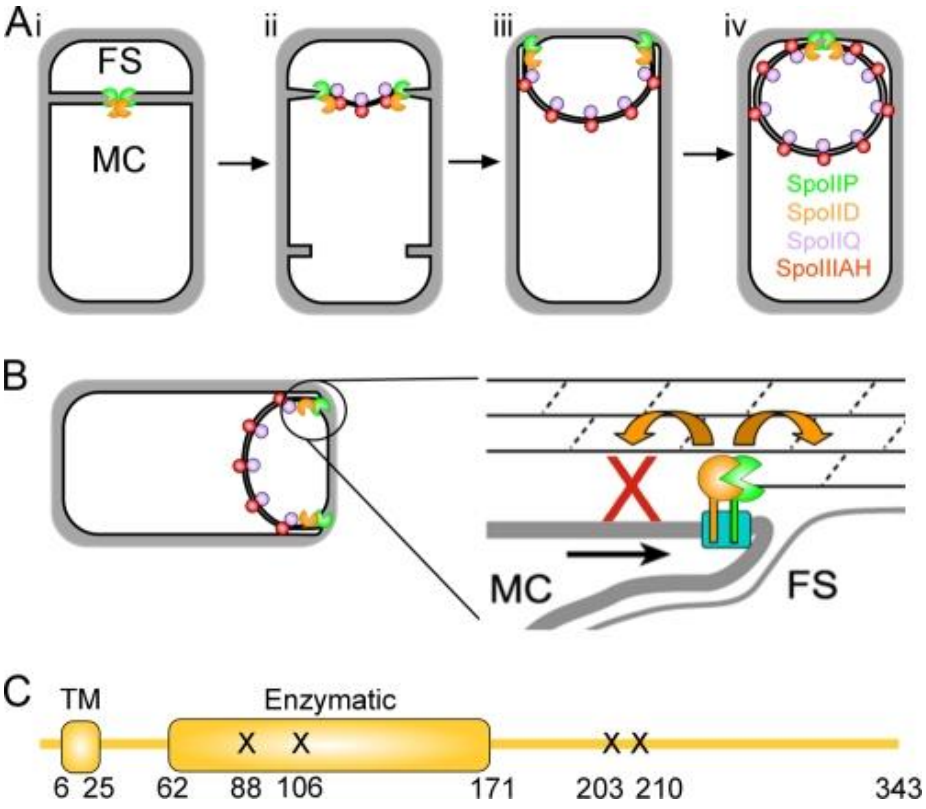

Fig 1. Action of the SpoIIPD complex in sporulation (A), a close view of the activity (B), conserved domains of the SpoIID (C). Taken with permission from (Gutierrez et al., 2010)

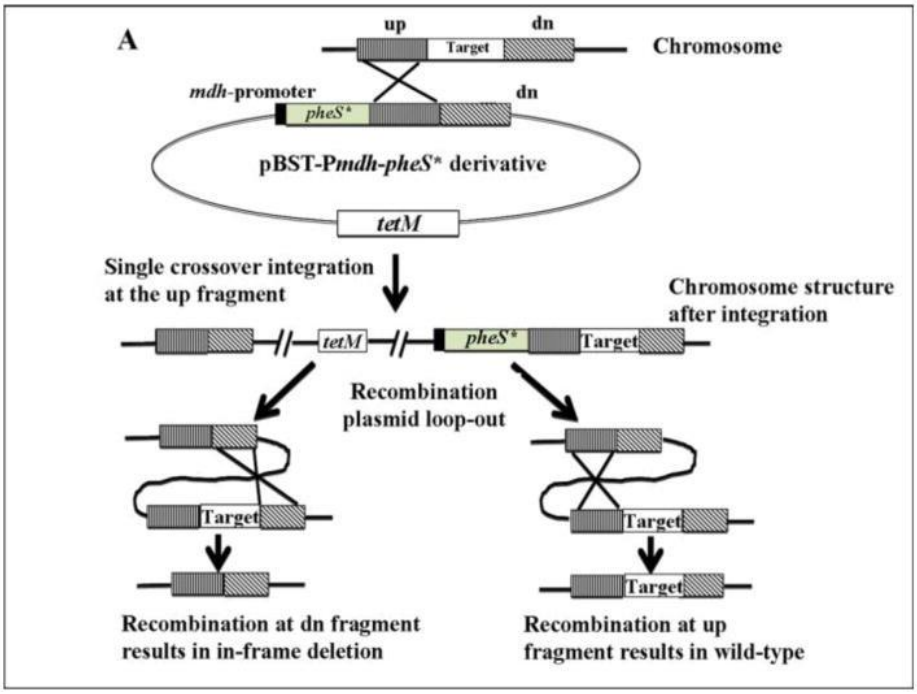

Fig 2. Schematic presentation of the strategy for constructing the markerless deletion system. Here only integration at the upstream region is illustrated. Integration can also happen at an equal chance in the downstream region. When this happens, the result from the second step is opposite to what illustrated here; i. e. recombination excision at the downstream region would recreate the wild-type genotype, while recombination at the upstream region would generate the deletion. Legend and figure took with permission from (Zhou et al., 2015)

Results

To mutate *spoIID* two distinct vectors were created (Figure 3), one for insertional mutagenesis, pBST-PmdH- spoIID-INS and another for in-frame deletion, pBST-PmdH-spoIID- pheS*-DEL. A third plasmid, pBSJL2, which was an empty expression vector, was used as a control.

For each of the four batches of competent cells 3 were electroporated with pBST-PmdH- spoIID-INS, 3 with pBST-PmdH PheS*--spoIID-DEL, 3 with the positive control pBSJL2, and 3 as a negative control that were not provided with a plasmid. These cells were allowed to recover and then 50% of each sample were plated on THL containing the antibiotic and 50% on THL without antibiotics.

I had near identical results in each of the four batches of competent cells with both deletion strains producing no living clones, while all of the Positive controls produced from 10-50 colonies per plate.

Discussion

The mutation of *V. parvula* could have failed for three reasons: Strain selection, low competence efficiency, or *spoIID* is an essential gene.

V. parvula DSM 2008 may have been a poor choice for performing this study. We chose this bacterium as it is the type strain for the Family and we had previously performed an OM proteome and knew SpoIID was present and expressed. Furthermore, we had evidence that we could obtain a single cross over event from previous analysis within the laboratory (Data not shown). Furthermore, the transformation efficiency is incredibly low, with an average of two colonies per µg of plasmid. The only successful mutations published (Liu *et al.*, 2012) have been done with a different strain, that of *V. atypica* OK5, which now has a genome available (Zhou *et al.*, 2017).

Another possibility is that the mutation was possible, however, there were too few competent bacteria to obtain a mutant. In our process we used electroporation for DNA uptake, however, a recent study has demonstrated that *Veillonella* species may use their T4P for competent uptake of DNA at higher efficiency than that of electroporation (Knapp *et al.*, 2017). I would recommend using the natural system whenever possible for all future studies to ensure high numbers of competent cells.

There is a strong likelihood that *spoIID* is an essential gene in diderm bacteria. First and foremost is that if our hypothesis is correct, deletion of *spoIID* results in a monoderm bacteria. With the massive number of bacterial populations that exist in the natural environment we would expect the monoderm/diderm transition to be a common occurrence, instead we only find evidence for 9 potential transitions in known Bacteria, in particular, the Actinobacteria, Chlorflexi, and 4-7 transitions within the Firmicutes (Sutcliffe, 2010; Antunes *et al.*, 2016). Additional support for *spoIID* being essential comes from our survey of peptidoglycan remodelling enzymes, *spoIID* is found in all diderm Firmicutes without exception and is only found in multiple copies within the diderm Firmicutes

that sporulate. This contrasts the PG hydrolases which have sporadic distribution throughout the Firmicutes and vary widely in copy number. Even the genes responsible for the cytoplasmic biosynthesis of the PG precursors are found in a single copy number. In addition *spoIID* may not mutate due to the local genomic context; It is downstream from the holiday junction resolvases and upstream from the sec system and may be on an operon with both (Figure 3). The sec genes are essential in *E. coli* (Gerdes *et al.*, 2003) and mutation of the holiday junction resolvases can prevent homologous recombination in and of itself (Wyatt and West, 2014).

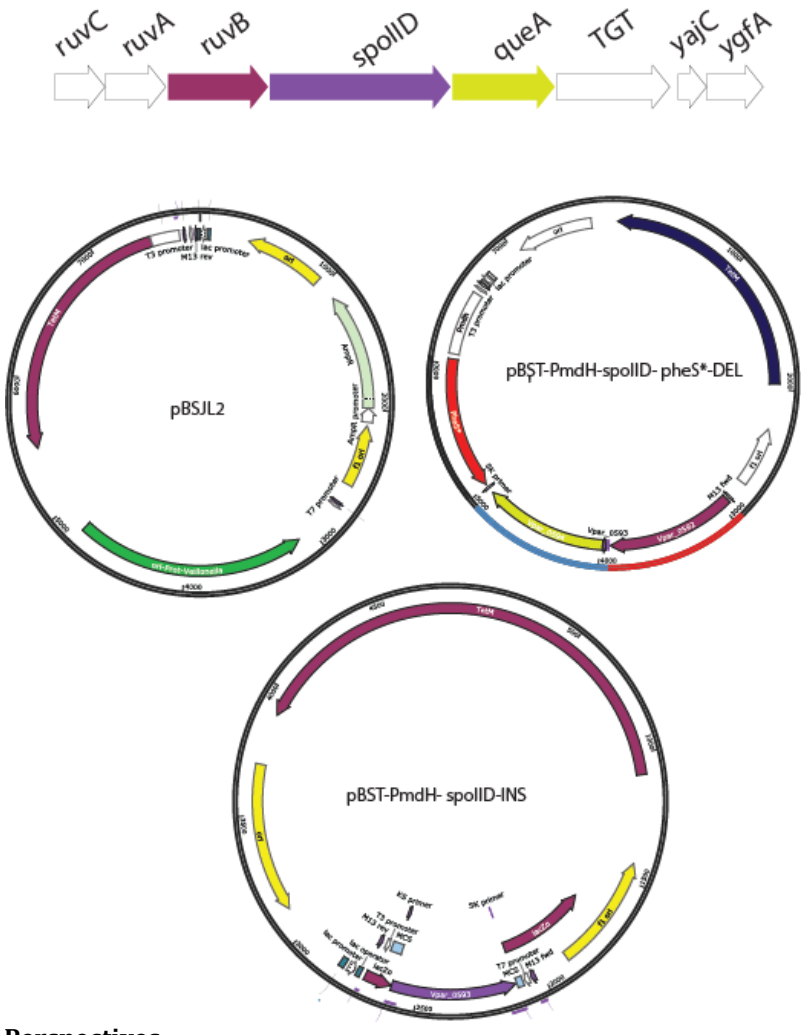

Fig 3. Local genomic area of spoIID (top) and plasmids used in this study (Bottom)

Perspectives

spoIID may be the gene responsible for the monoderm/diderm transition, however, it may not be possible to perform this mutation in *V. parvula* DSM 2008. For future studies, I would recommend attempting to express a *Veillonella* type *spoIID* in *B. subtilis* during sporulation and see if the OM is retained. Should mutation be attempted I would suggest using the strain that has the most successful mutations, namely *V. atypica* OK5 (Zhou *et al.*, 2017) and attempt overexpression in parallel.

Materials and Methods

The growth of competent cells. *V. parvula* DSM 2008 was grown in Todd–Hewitt (TH) broth (Difco) supplemented with 0.6% sodium lactate (THL) for 24 hours. This was subcultured 1:100 into 200 mL and grown for 8 hrs @ 37ºC to an OD of 0.15. The cells were harvested by centrifugation at 6000 rpm for 10 min @ 4ºC and resuspended in 100 ml of EpB (10 % glycerol, 0.5 mM MgCl¬2). Cells were harvested using the same speed and temperature and then resuspended 25 ml of EpB. Cells were harvested using the same speed and temperature and then resuspended 1 ml of EpB. Cells were aliquoted into 80 µl fractions and froze immediately. Cells were stored at -80ºC for less than one week.

Electroporation. 10 µg of pure plasmid was combined with one 80 µl aliquot of cells and electroporated in a 1 mm cuvette with 20 kV*cm-1, 25 µF, and 200 ohms. Cells were diluted into 5 ml of THL and grown anaerobically for 16 hrs @ 37ºC. 2 ml of the cells were plated on THL and additional 2 ml were plated on BHIL+TET (2.5µg/ml). Plates were grown anaerobically for at least 72 hours @ 37ºC.

Antunes, L.C., Poppleton, D., Klingl, A., Dupuy, B., Brochier-Armanetd, C., Beloin, C., and Gribaldo, S. (2016) Phylogenomic analysis supports the ancestral presence of LPS outer membranes in the Firmicutes. Elife e14589 https://elifesciences.org/content/5/e14589.
Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balázsi, G., Ravasz, E., Daugherty, M.D., et al. (2003) Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. J Bacteriol 185: 5673–84 http://www.ncbi.nlm.nih.gov/pubmed/13129938. Accessed August 1, 2017.
Gutierrez, J., Smith, R., and Pogliano, K. (2010) SpoIID-mediated peptidoglycan degradation is required throughout engulfment during Bacillus subtilis sporulation. J Bacteriol 192: 3174–86 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2901704&tool=pmcentrez&rendertype=abstract. Accessed September 16, 2015.
Knapp, S., Brodal, C., Peterson, J., Qi, F., Kreth, J., and Merritt, J. (2017) Natural Competence Is Common among Clinical Isolates of Veillonella parvula and Is Useful for Genetic Manipulation of This Key Member of the Oral Microbiome. Front Cell Infect Microbiol 7: 139 http://www.ncbi.nlm.nih.gov/pubmed/28473967. Accessed May 18, 2017.
Liu, J., Xie, Z., Merritt, J., and Qi, F. (2012) Establishment of a tractable genetic transformation system in Veillonella spp. Appl Environ Microbiol 78: 3488–91 http://www.ncbi.nlm.nih.gov/pubmed/22344660. Accessed August 1, 2017.
Nocadello, S., Minasov, G., Shuvalova, L.S., Dubrovska, I., Sabini, E., and Anderson, W.F. (2016) Crystal Structures of the SpoIID Lytic Transglycosylases Essential for Bacterial Sporulation. J Biol Chem 291: 14915–14926 http://www.ncbi.nlm.nih.gov/pubmed/27226615. Accessed August 1, 2017.
Sutcliffe, I.C. (2010) A phylum level perspective on bacterial cell envelope architecture. Trends Microbiol 18: 464–470.
Wyatt, H.D.M., and West, S.C. (2014) Holliday junction resolvases. Cold Spring Harb Perspect Biol 6: a023192 http://www.ncbi.nlm.nih.gov/pubmed/25183833. Accessed August 1, 2017.
Zhou, P., Li, X., and Qi, F. (2015) Establishment of a counter-selectable markerless mutagenesis system in Veillonella atypica. J Microbiol Methods 112: 70–2 http://www.ncbi.nlm.nih.gov/pubmed/25771833. Accessed August 1, 2017.
Zhou, P., Xie, G., Li, X., Liu, J., and Qi, F. (2017) Complete Genome Sequence of Veillonella atypica OK5, the First Transformable Strain in the Species. Genome Announc 5: e00391-17 http://www.ncbi.nlm.nih.gov/pubmed/28572307. Accessed August 1, 2017.

CONCLUSIONS AND PERSPECTIVES

We are in the era of microbial genomics. The field is exploding with new genomes, representing the diversity that was only imagined before. When this doctoral work began there were genomes available for 38 Negativicutes and 7 Halanaerobiales; now there are 83 Negativicute and 18 Halanaerobiales, a doubling of available genomes. This expansion in available genomes is not limited to Firmicutes or known strains, as there have been huge leaps in bacteria characterized from metagenomes and new environments with 73,000 out of 100,000 prokaryote genomes being released during the course of this thesis (2014-2017, NCBI prokaryotes.txt). A paradigm shift is necessary as an estimated 2% of bacteria are culturable and of those less than 50% are estimated to be isolated at the species level (Wade, 2002). Through sequencing, we are able to obtain new perspectives on these uncultured, difficult to study organisms. This comes with a problem as we have increased the amount of data exponentially, however, the number of researchers has remained relatively constant. This work represents a consolidated approach to this explosion of data and the generation of new sequence annotation and genomes.

In this thesis, I used the available sequence data to construct an evolutionary story unlike that told before. I increased the available knowledge on the Firmicutes, in both annotation of OM related genes and describing the evolutionary relationship (Results Section 1). In addition, I have added to the available knowledge by binning three Limnchorda from metagenomes and describing their novel OM systems (Results Section 1.2). These results have provided evidence for a unique theory that is a diderm first hypothesis. Our results present a diderm ancestor of the Firmicutes who lost the OM in multiple lineages. This contrasts the Cavalier-Smith diderm first hypothesis (Cavalier-Smith, 2006a), as we have demonstrated multiple independent losses in the Firmicutes alone compared to his declaration of a single loss in Bacteria. Furthermore, we discredit many of the monoderm first hypothesis, most especially the Gupta hypothesis (Gupta, 2011). First, we show the loss of the OM, not the gain, and we have shown the presence of LPS producing enzymes in multiple phyla with no evidence of transfer (Results Section 1.2). This disproves the development of an OM multiple times independently.

This work is on the forefront of science due to the combination of *in silico* and experimental work. Science has gone through transitions of philosophy, methodology, and people. One of the current transitions is from two distinct methods of study, bioinformatics and laboratory based study, to a holistic approach that encompasses both methodologies. This approach is exemplified in the OM proteome (Results section 2.1). In this section, we managed to definitely demonstrate the proteomic characteristics of *V. parvula*'s OM. Through this, we were able to characterize ancestral systems such as BAM/TAM and we uncovered multiple new hypothetical proteins, which will be a priority for future studies. The quantification from this study directly led to the OM-tethering question; As Lpp is the most abundant protein in E. coli and the functional similar OmpM is the most abundant protein in *V. parvula*, what mechanism do other bacteria use? To this end I combined a literature review and a thorough bioinformatics search for OM attachments and broke the textbook description (Results section 2.2); Lpp is drastically limited in distribution and other mechanisms dominate. These other mechanisms should be explored in more depth to determine their role in OM-tethering outside of bacteria with Lpp. Furthermore, mutational studies on the diderm Firmicutes would be informative to determine how the system function.

Although this work did not definitively demonstrate the mechanism of OM loss in the monoderm Firmicutes, I was able to produce a strong logical hypothesis on the basis of SpoIID (Results section 3). This is the first identification of a single enzyme that may be responsible for a drastic shift in cell envelope architecture. Unfortunately, the mutagenesis failed in *V. parvula* and this demonstrated the need for more developed techniques in the diderm Firmicutes. The only diderm Firmicute with potential mutagenesis is *V. atypica*, however, this needs to be extended out to other Negativicutes, especially *Selenomonas ruminantium,* and other diderm Firmicutes. The Halanaerobiales have much to teach us as they live at the limits of life and being able to study their genetics will extend the knowledge on bacterial extremophiles and their unique enzymatic and physiological properties. With genetic systems in the diderm Firmicutes we could explore the nature of the unique proteins that we have identified. These have been discussed in Results section 1.1, 1.2 and, 2.1, however, emphasis should be placed on the hypothetical upstream of the BAM system. This protein may give strong insight into how diderm Firmicutes envelope functions. The many other hypotheticals discussed in this work could have potential as well and a strong deletion system such as transposon mutagenesis need development.

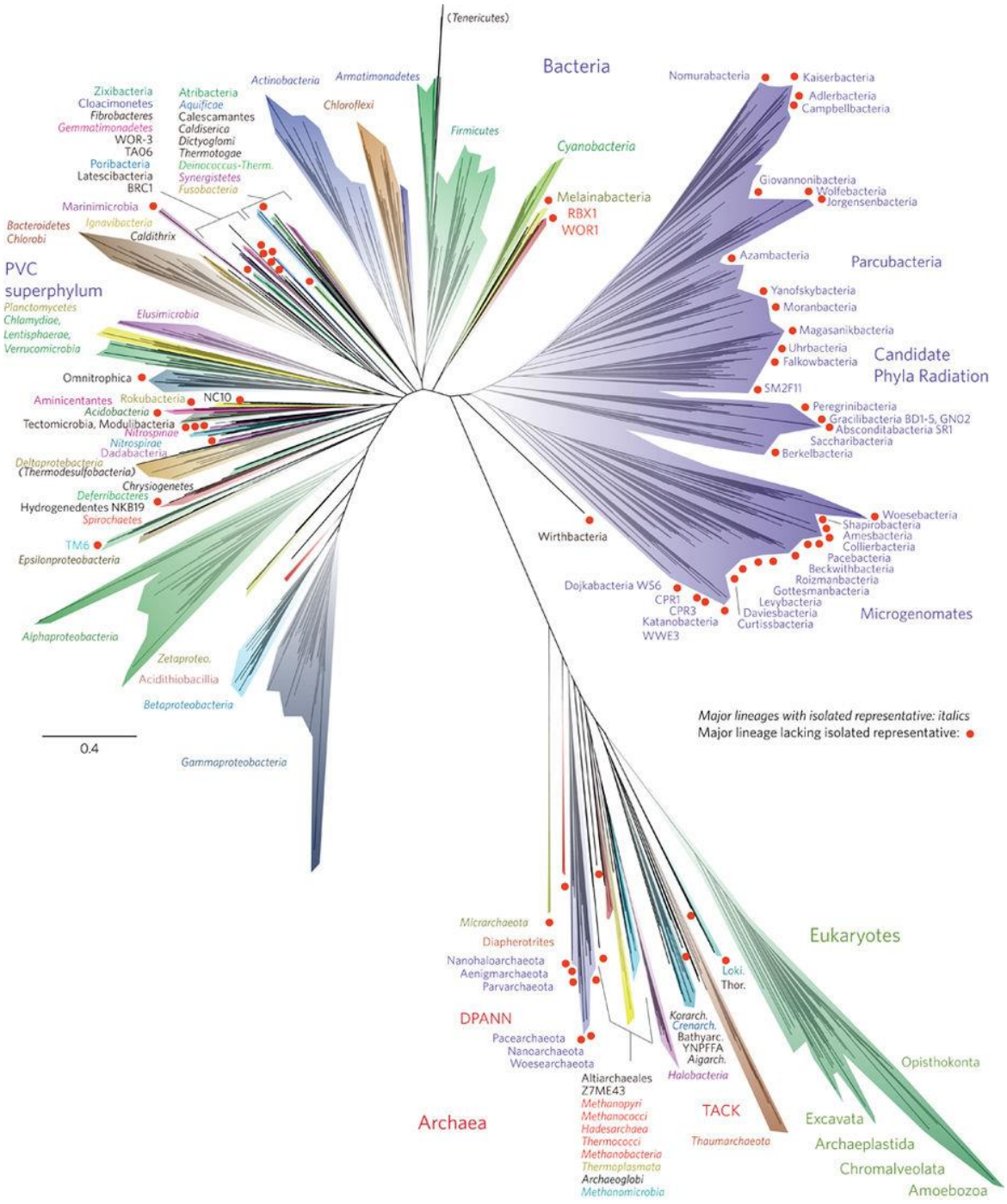

Figure D.1. New Views on the tree of life.

The tree includes 92 named bacterial phyla, 26 archaeal phyla and all five of the Eukaryotic supergroups. Major lineages are assigned arbitrary colours and named, with well-characterized lineage names, in italics. Lineages lacking an isolated representative are highlighted with non-italicized names and red dots. For details on taxon sampling and tree inference, see Methods. The names Tenericutes and Thermodesulfobacteria are bracketed to indicate that these lineages branch within the Firmicutes and the Deltaproteobacteria, respectively. Eukaryotic supergroups are noted, but not otherwise delineated due to the low resolution of these lineages. The CPR phyla are assigned a single colour as they are composed entirely of organisms without isolated representatives, and are still in the process of definition at lower taxonomic levels. Taken with permission from (Hug et al., 2016).

Not only do we need more genetic systems and studies we need more genomes of diderm Firmicutes. The predominantly studied diderm Firmicutes are associated with mammalian microbiomes, a very well-studied approach (Turnbaugh et al., 2007). We need to explore more environments that may contain interesting diderm Firmicutes. To this end, the metagenomics work that I have performed here (Results section 1.2) should be extended to the other diderm Firmicutes. In particular, the Halanaerobiales were recently found in fracking water (Daly et al., 2016) and other locations could have other Halanaerobiales. Similarly, the Negativicute genus *Pelosinus* has been found in multiple Uranium contaminated sites. Continuing to search through metagenomics samples and generating novel data from untouched environments could increase our knowledge on the diderm Firmicutes and may enable the discovery of more lineages that represent distinct diderm clades.

In the broader approach, the work on cell envelope architecture has been greatly expanded in the recent years. I have summarized this data in an updated version of Sutcliffe's table (Table D.1). The diversity of Bacterial phyla is incredible as we currently have 30 described phyla in Bergey's manual (Whitman, 2011), however, there is probably 92 phyla found within available genomic data (Hug et al., 2016). Only 36 of these probable phyla have an isolated representative and definitive proof of their envelope architecture (Figure D.1). Four well established monoderm clades are present within Bacteria: Firmicutes, Tenericutes, Chloroflexi, and the Actinobacteria. The Tenericutes have been shown to be placed within the Firmicutes (Wolf et al., 2004) and as such, they lost their OM like all other monoderm Firmicutes. The CPR have been proposed to constitute multiple phyla and have many unique characteristics. All are extremely small with reduced genomes and most important for this discussion is that they contain both diderm and monoderm members. Although their phylogeny is difficult due to a reduced genome and fast evolving members, it has been suggested that this group is a basal phylum. Repeating the same analysis, we performed on the Firmicutes with the CPR may enlighten us on the ancestral Bacterial envelope and whether it was diderm or monoderm. Unfortunately, this group has no cultivated members and only partially completed genomes. When these problems have remedied a repeat of this analysis would be ideal to be performed and may let us know what the LCA of bacteria possessed for a cell envelope.

The Firmicutes have inherited an OM and lost this envelope multiple times within the phylum, however, the Actinobacteria have an opposing story. Within this phylum diderm members exist, the Corynebacterineae (Kaur et al., 2009; Bansal-Mutalik and Nikaido, 2014),

which possess an OM. The Corynebacterineae are a well-known clade as Mycobacterium tuberculosis is a member. Unlike the Firmicutes the origin of this OM is not a complete mystery as it has emerged over the development of this order (Jamet et al., 2015). This OM is non-homologous to any other known diderm bacteria and consists of a layer of Arabinogalactan covalently attached to the exterior of the PG with hydrophobic mycolic acids on the exterior ends of the mycolic acids (Kieser and Rubin, 2014). This unique case of convergent evolution is the only time *de nova* membrane biosynthesis has been observed since the origin of life. As this group contains the Mycobacteria many different evolutionary studies have been performed, however very few have focused on the origin and emergence of this OM. Although it would be interesting to study the development of an OM, we can extend beyond and ask did LUCA have an OM? I will address my thoughts on this question in the opinion section.

Phylum	Observable OM	Characterized LPS	LPS Genes Detected	Characterized BAM	BAM/TAM genes detected
Acidobacteria	Present	Present	Present	-	Present
Actinobacteria	Present^	-	Absent	-	Absent
Aquificae	Present	Present	Present	-	Present
Armatimonadetes	Present (Tamaki *et al.*, 2011)	-	Present	-	Present (NP)
Bacteroidetes	Present	Present	Present	Present	Present
Caldiserica	-	-	Absent	-	Absent (NP)
Chlamydiae	Present	Present	Present	-	Present
Chlorobi	Present	Present	Present	Present	Present
Chloroflexi	Absent (Sutcliffe, 2011)	-	Absent	-	Absent*
Chrysiogenetes	Present (Rauschenbach *et al.*, 2011)	-	Present	-	Present (NP)
Cyanobacteria	Present	Present**	Present**	Present	Present
Deferribacteres	Present	-	Present	-	Present (NP)
Deinococcus-Thermus	Present	-	Present**^	Present	Present
Dictyoglomi	Present	-	Present**	-	Present (NP)
Elusimicrobia	Present	-	Present	-	Present (NP)
Fibrobacteres	Present	-	Present**	-	Present (NP)
Firmicutes	Present^	Present**^	Present**^	-	Present^
Fusobacteria	Present	Present	Present	-	Present
Gemmatimonadetes	Present	-	Present	-	Present (NP)
Lentisphaerae	Present (Limam *et al.*, 2010)	-	Present**	-	Present (NP)
Nitrospira	Present	Present	Present	-	Present (NP)
Planctomycetes	Present	Present	Present	-	Present
Proteobacteria	Present	Present	Present	Present	Present
Spirochaetes	Present	Present^	Present^	Present	Present
Synergistetes	Present (Jumas-Bilak *et al.*, 2009)	-	Present	-	Present (NP)
Tenericutes	Absent	-	Absent	-	Absent
Thermodesulfobacteria	Present	-	Present	-	Present (NP)
Thermotogae	Present	-	Absent	-	Present (NP)
Verrucomicrobia	Present	Present	Present	-	Present

*Some homologs found **Key gene absences ^Not found in all clades of phylum
- Indicates unknown (NP) Not published

Table D1. Review of Bacterial Cell Envelope Architecture.
The thirty described phyla are from Bergeys (Euzeby, 1997), and probable diderm groups are in bold. Observable OM references are from (Whitman et al., 2011.) or are included in the table. Characterized LPS is taken from (Sutcliffe, 2010) while LPS genetic potential is taken from (Antunes et al., 2016). Published data relating to BAM/TAM system is reviewed in or taken from (Chaille T Webb et al., 2012; Heinz et al., 2015)

Opinion:

The Origin of the Outer Membrane and Life Itself

During the course of my thesis the fundamental question of “Was the last universal common ancestor a diderm or monoderm?” has plagued my thoughts. In this opinion section, I present my ideas on where it began and what I believe happened.

Current Opinion by
Daniel

Starting Outside the Box: Origin of Life on a Ball

Daniel Poppleton
Unité de Biologie Moléculaire du Gène chez les Extrêmophiles, Département de Microbiologie, Institut Pasteur, Paris, France

In this opinion piece, the origins of life are described. First abiotic synthesis of molecules occurred, these grew in complexity and aggregated. The molecular aggregates or bioballs underwent selection by uncontrolled division and selective accumulation of molecules. Then, more complex equipment evolved by the evolution of molecules on the outside of the bioball, which started to develop a cavity. The bioball invaginates and the protocell continues development past LUCA. Then the invagination closes and the LCA of Archea and Eukaryotes while Bacteria maintained the OM and pore until they had developed peptidoglycan and the complex transport machinery.

Step 1: Warm Little Pond

Darwin once stated in a letter to J. D. Hooker "It is often said that all the conditions for the first production of a living organism are now present, which could ever have been present. But if (& oh what a big if) we could conceive in some warm little pond with all sorts of ammonia & phosphoric salts, light, heat, electricity & present, that a protein compound was chemically formed, ready to undergo still more complex changes, at the present-day such matter be instantly devoured, or absorbed, which would not have been the case before living creatures were formed. (Darwin and Darwin, 1888)" Darwin's novel idea was brilliant and since then many different mechanisms have been proposed for the chemical synthesis of the original biomolecules of life (For a recent review see (Ruiz-Mirazo *et al.*, 2014)). A common feature of proposed prebiotic reactions, and chemistry in general, is heterogeneous production. Even in a controlled laboratory reaction, a yield of 70% of the desired product is considered good. Alas, most current researchers on the topic seem to focus on the production and implications of a single molecule such as RNA (Materese *et al.*, 2017) or fatty acids (Mansy, 2010). It is impossible that the early reactions produced only one product, such as glycine; Far more likely is the prospect of a heterogeneous mixture of organic molecules, from lipids to sugars. There are many examples of messy prebiotic reactions (Schrum *et al.*, 2010; Ruiz-Mirazo *et al.*, 2014) as a specific example I will use the basics of Zhang *et al* in their paper on Mineral assisted pathways (Zhang *et al.*, 2004). In this groundbreaking work, they present data that suggests that photochemical reactions combined with mineral colloids can produce small organic building blocks and longer chain organic molecules. From these experiments I will make a drastic assumption: **Life started in a heterogeneous environment with a wide range of organic molecules**, ranging in size, charge, and hydrophobicity. Although incredibly interesting, I will make no assumptions as to how these molecules are produced or to the location of their production as this is not the scope of this chapter. I will assume that it occurred in an aqueous environment, with a near neutral pH, at temperatures similar to those promote life today (5-80ºC).

.

Step 2: Nucleation

Now small molecules react, but for the synthesis of life, we need large molecules, such as saccharides and peptides. Here we run into a problem, the formation of large molecules is difficult as they are more likely to split than elongate. This is due to the fact that bond formation is the product of second-order kinetics, while breakage is first order. To overcome the hydrolysis hurdle, the concentration of the reactants has to be incredibly high, either by high production (unlikely) or by increasing the working concentration. The working concentration can be increased by the production of crowding agents, such as polyethylene glycol. This likely contributed in a messy environment, but could not raise the concentration by the amount needed for the speedy reaction time necessary (Houston, 2006). I propose an alternative mean: Nucleation.

By nucleation, I propose that **the hydrophobic effect would cause the accumulation of hydrophobic molecules into a single mass separate from that of water**. To form this hydrophobic interface the free energy must increase. Logically the interface cost grows linearly with surface area, while the opposing compensation forces grow linearly with volume. Therefore this cluster can only be stable if it exceeds a certain size

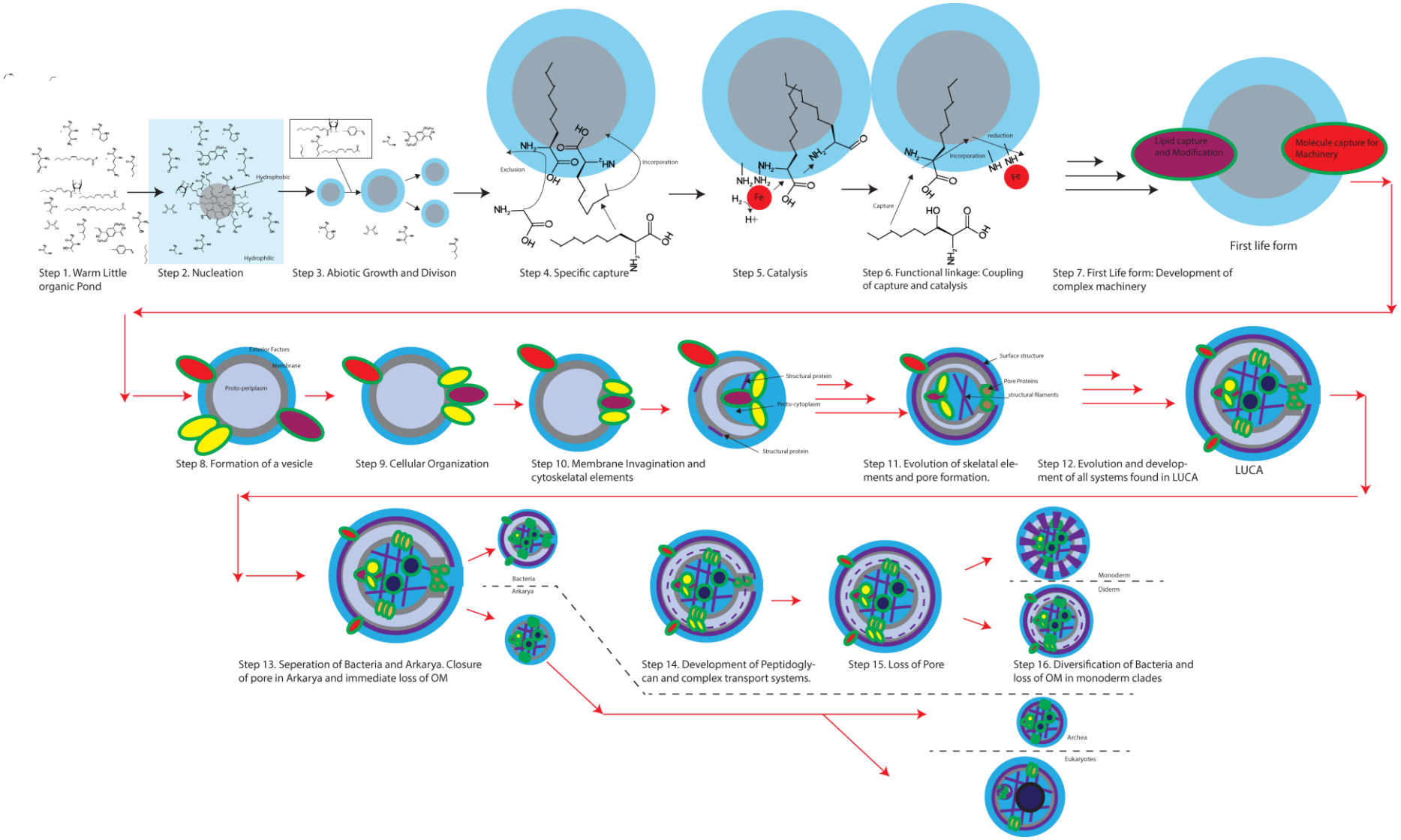

Figure 1. The Development of Life

For a detailed description see the text of this chapter. Blue represents a hydrophilic environment while grey is a hydrophobic region or membrane. Large black arrows note development by random chance, while red arrows denote development by selection. Mechanical systems have green borders and structural elements are in purple.

in this case 1nm (Dwivedi *et al.*, 2014). This is similar to nucleation of crystals, which can only phase separate after the formation of a seed. 1 nm is not a problem size. Linear glucose is 1.5 nm and as such a single five carbon hydrocarbon could act as a nucleation seed.

In this messy reaction scheme a single hydrophobic molecule is uninteresting, however, a mass of amphipathic molecules is. Take one of our previous assumptions, that this was an incredibly heterogeneous mixture of organic molecules, all of the molecules that are hydrophobic will come to together favourable to form a mass *and so will the amphiphatic ones.* This may seem obvious from vesicle formation, where polar head groups on hydrophobic tails form a sphere, but most theorists assume a mass of homogeneous content with a single type of molecule constituting the vesicle (Kurihara *et al.*, 2011; Hentrich and Szostak, 2014). I favour the idea that these organic balls would be incredibly heterogeneous, containing both fully saturated hydrocarbons in its core and amphipathic molecules near the outside. This mass of molecules, which I shall call a bioball, has key properties that make it favourable to produce life: catalysis, growth and division.

Step 3. Abiotic Growth and Division

The first property of the bioball is that it can grow and divide. The growth aspect is easy to comprehend; these bioballs will incorporate whatever is thermodynamically favourable, both hydrophobic and hydrophilic molecules. As this will be a completely random process, many of these incorporations may not decrease the free energy of the system, as such hydrophobic portions may be exposed if too many hydrocarbons are present in the core. Furthermore, the shape of the bio-ball is determined by the surface tension equation, which specifies that the tension of the surface inversely proportional to the cube of the volume. Briefly, this means that as the ball grows, it is more likely to lose its shape. **This loss of shape will cause a division into two separate bioballs when the bioball acquires enough mass**. It may seem like we are on the path to life at this point as we have a replicating entity, however, there is no selection or synthesis of components.

Step 4. Specific Capture

At this point, the abiotic replicating bioball is a complete heterogeneous mess, composed of a random assortment of molecules. A key phase in its development is the ability to preferentially grab and incorporate molecules. Take, for instance, a 10-carbon molecule, of which 5 carbon atoms are reduced and hydrophobic, while the other 5 are negatively charged. This molecule would preferentially grab positively charged molecules from the environment. If these "adhesin" molecules have a hydrophobic portion they will be incorporated into the bio-ball, otherwise, they will eventually disperse. Then this incorporated molecule would preferentially grab positively charged molecules and the cycle would continue. Molecules that bind poorly to any exposed molecule would be repelled from the bio-ball and this heterogeneous system would develop into a more homogeneous molecule. This process would be expanded by messy capture, where a molecule would preferentially grab a copy of itself, but also a second molecule that may have another function. **If you extrapolate this idea to more complex molecules, we can have capture complex molecules similar to base pairing in DNA**. These exposed molecules would be the earliest form of a genome, providing the rudimentary code for incorporation of molecules.

Step 5. Catalysis

A difficult property of these early reactions is concentration, as stated before; most organic synthesis reactions are second or third order. If we take the assumption that longer chain molecules will be more abundant than smaller molecules, large molecules will interact with the di and triatomic molecules before other long chain molecules. Should some object hold the substrates in close proximity, such as a bioball, the reaction order can be decreased and the rate-limiting step will be the synthesis reaction rather than the molecules finding each other in the primordial soup. This means that the structure of the bioball, in and of itself, would act as a catalyst, but it would not be the only one.

In our messy environment, some organic molecules will be produced, by random chance, that exhibit catalytic activity. Organocatalysis is the field of chemistry that focuses on catalysis using small organic molecules, such as those that may have to be produced in the prebiotic environment. These molecules can catalyse most organic synthesis reactions; for example, proline can catalyse addition reactions (Yao and Yuan, 2013) and a frustrated Lewis pair can catalyse hydrogenation (Rochette *et al.*, 2015). I do not want to say all early chemistry was catalysed by proline, this is simply a proof of concept until we know the exact reaction conditions, we will not what acted as "enzymes". Organocatalysis may have been one branch of early reactions, however, an equally likely form would have been organometallics. This bioball would have many exposed amine and carboxyl groups, the same groups that can chelate iron, manganese, and other metals commonly used in biochemistry today. These organometallic complexes could perform many of the necessary reactions without much-added complexity such as

reduction such as nickel reduction of carbon. **This small molecule catalysis would be the essence of early non-protein enzymes.**

Step 6. Functional linkage: Coupling of capture and catalysis

Now let us extend this idea by logical progression. By random chance, we may have one of the adhesion molecules that is incorporated in the same bioball as a catalyst. This may allow capture of a molecule and specific modification of it. For example, if you had a dehydrogenation paired with a capture mechanism, we may have the first lipid synthesis system. These systems would exist in isolation, aiding the bioball they are in, yet unable to replicate to both daughter cells. A drastic development would be the coupling of such systems to produce a self-replicating system.

Step 7. First Life form: Development of complex machinery

At this point, we have the first glimpse of life. We have a **replicating entity that grows in a preferential manner, incorporating specific molecules and modifying them** with the ancestor of metabolism. The genetic code is in the functional molecules it exposes to the environment. Until this time, the bio-ball developed purely by kinetics and random chance. Now selection can act on this proto-lifeform and I would argue that this is the first living particle. The first self-replicating bioball would cause depletion of its primary substrates in the environment. This depletion facilitates diversion and "mutation". When the self-replicating machinery has no substrates the bio-ball would be susceptible to decay and may incorporate molecules other than its preferred substrate. These alternative substrates may confer a different activity or be disastrous for the bio-ball, similar to what we see in genomic evolution today.

Inside Out Metabolism

The process of cellular system development would continue for an incredible amount of time and all cellular systems would develop exposed to the environment, as opposed to the interior. I will avoid discussion on the development of individual systems such as polymerases, RNA world, as the bulk of the research in this field has been dedicated to these systems (Wu *et al.*, 2009; Ruiz-Mirazo *et al.*, 2014) What I wish to propose is the expansion of the bio-ball idea to LUCA in terms of cellular organization.

In this chapter, I am following and developing the ideas first put forth by Blobel (Blobel, 1980) and later developed by Cavalier-Smith (Cavalier-Smith, 1987; Cavalier-Smith, 2006a). Blobel first proposed the idea that the ancestor of all life was a double membrane organism and lineages later lost the OM. This work was extensively expanded by Cavalier-Smith with detail on mechanism and acquisition of systems. I will not expand on Cavalier-Smith's work on individual mechanisms but highlight the key differences. On this topic, most modern research suggests that LUCA (Last Universal Common Ancestor) developed as a cell (For an apt summary of multiple viewpoints please see (Koonin, 2014). This work was enhanced by Gunter Wachtershauser in 1988 (Wächtershäuser, 1988), however, he presented evidence for development on an inorganic surface, while I propose the development on an organic surface, the bio-ball. This development would continue far past the usual considerations of an early cell and right up to LUCA and all cellular aspects would evolve on the surface of the bioball. These aspects include all of LUCA's features including ribosome biogenesis, Nucleic acid biosynthesis, catabolism of major substrates, anabolism of amino acids and sugars, and transport mechanisms.

As this is a controversial area, I give my reasons and logic here.

Current opinions are that the biochemical systems we know today (RNA polymerase, Ribosome, etc) would develop in a cellular environment. I disagree with this hypothesis and follow the logic of surface metabolism for the following reasons.

Membranes are never synthesized de nova. As of yet, no one has found the *de nova* synthesis of a membrane, all membranes in existence are inherited from the parent cells (Robertson 1964). Membranes grow by the incorporation of new molecules and cells grow by division and budding from an existing structure. Even in the case of sporulation, where a complex second membrane is formed, the membrane is formed from the cytoplasmic membrane, not *de nova*. I cannot exclude the possibility that *de nova* synthesis existed in the past, however, if it did; Why wasn't such a useful mechanism kept in some extant organisms? The only exception is the Mycobacterium, whose membrane has a truly unique structure.

Membranes inhibit diffusion. The main advantage of a membrane is having a semipermeable separation from the exterior environment of the cell. All cellular membranes allow passage of small molecules, but the speed of diffusion increases exponentially with size. This means only small molecule chemistry could have existed in the early days of life, an unlikely proposition.

Cart before the horse. As only small molecules, at most a single sugar, can readily pass the membrane we require a transporter or pore and secretion system before life can develop. These systems require biosynthesis to be specific enough to be beneficial. If they are a non-specific pore there is no advantage to having metabolism on the inside. As

such the complex transport systems, such as SEC, require protein biosynthesis, protein biosynthesis requires ribosomes, and ribosomes required coding material. This means that we must have complex mechanical machinery before life can be cellular hence the horse must come before the cart.

Cellular life could not compete. As I stated before early life developed in an environment that favoured organic molecule synthesis. This initial synthesis must have resulted in at least a self-replicating molecule, which would be acted on by selection. If this molecule is separated from its substrates it would have reduced fitness compared to the surface metabolism counterparts.

Membrane loss is found in Bacteria. The bulk of this thesis has focused on the loss of an OM in the Firmicutes. As such we have no evidence of membrane acquisition, but we have evidence for membrane loss in one model case. This follows the evolutionary theory that evolution is dictated by gradual change and specialization until a loss confers an advantage at which point a drastic diversification is enabled. If you look at terrestrial Bacteria, 50% of the isolates are monoderm bacteria which only accounts for 5% of the bacterial phyla (excluding the newly discovered CPR). The same could have happened multiple times in history.

Step 8. Formation of a vesicle

Now the bioball possesses saturated hydrocarbons at its core. This feature would stay until the synthesis or incorporation of fatty acids. Once a typical fatty acid is preferentially synthesized the ball can develop a cavity within it naturally (Hentrich and Szostak, 2014). This cavity would have little use in the initial stages, as all membranes inhibit the transfer of molecules and all catalysis activity occur on the surface of this pre-cell. This membrane could not have the mechanical strength to hold structure, which would enhance division in early stages, later stages structure would come from surface proteins. This division could be facilitated by crimping of the cell with proteins similar to that of FtsZ in Bacteria. These systems would have many features at this stage that are completely absent from modern life. Almost all molecules would require some form of lipid attachment and even coding machinery would have a hydrophobic domain within the molecule.

Step 9-12 Development of cellular systems

A key point in the development of the proto-cell is asymmetry. Once surface organization and specialized regions have ripened, this starts the formation of an invagination. The surface machinery would interact strongly with each other to both promote catalysis and efficient capture of new molecules. Should these molecules replicate in a curved fashion, they could cause invagination of the cellular structure. At first, this invagination would allow the hydrophilic interactions of distal components and slow the diffusion of small molecules, later this invagination would create a new cavity, a proto-cytoplasm. This proto-cytoplasm is where most of the cellular development would occur, from DNA to ribosomes. The proto-cytoplasm would extend, aided by cytoskeletal factors within the proto-cytoplasm and exterior exoskeleton structure provided by a layer similar to an S-layer. When the external environment changed significantly from what would be favourable for cellular life the cavity would become an enclosed environment by the formation of a large controlled pore to facilitate transport of substrates into the cell. This topic has been extensively covered by Cavalier-Smith (Cavalier-Smith, 2001) and I do not wish to dwell on these features.

Step 13-15 Continued development of cellular systems

This proto-cell develops for an extended period. All aspects found in LUCA are developed in this proto-cell and the invagination pore is not closed until *after* the separation of the Arkarya (Group of both Eukaryotes and Archea) and the Bacteria. I differ on this point then Cavalier-Smith who proposed that Arkarya's ancestor is within the Bacteria. After my proposed separation the lineages separate due to the Arkarya losing the outer membrane becoming a true cell, while Bacteria maintain this structure and utilize it as a second compartment, the periplasm. The accounts for the vast differences in Bacteria and the Arkarya. This compartment does not close until the peptidoglycan structure has been developed and efficient transport mechanisms (ABC transporters, T2SS) have been developed and is no longer necessary.

Additional points: Mass extinction and Machinery loss

Why do we not see any bioballs in modern life? My hypothesis immediately suggests points of mass extinction. The early self-replicating life forms (Step 7) appear as nothing but food to any modern organism and would not exist past the point of catabolism. These unprotected bioballs would easily be digested by the later bioballs (Step 10-11) and disappear off the face of the earth. This same logic can be applied to LUCA and its extinct cousins. The loss of the OM in Arkaya gave a strong metabolic advantage; it was no longer necessary to synthesize many of the components found on the exterior. The alternative was Bacteria which developed the necessary components to compete with its simpler single membrane relatives.

A key point is that much of the machinery (pore proteins, reduction mechanism) may have conferred no advantage at key developmental stages including the LUCA bottleneck and were subsequently lost. What we see today is likely a small portion of the biologically active molecules that ever existed.

Where to go next

We are nearing the capabilities as a species to discover and understand one possibility of LUCA. Although we can never truly know what actually happened without Dr. Who, we can discover mechanism and maybe create life. I propose a simple combination of chemistry and computational modelling. In this experiment, we attempt to replicate each of the conditions hypothesized on early earth and record what the products of these reactions are. This data is fed into a combination neural-network/molecular simulation program. Continued building of this knowledge will inform us on what the earth looked like and where we may have come from.

Blobel, G. (1980) Intracellular protein topogenesis. Proc Natl Acad Sci U S A 77: 1496–500.

Cavalier-Smith, T. (1987) The origin of cells: a symbiosis between genes, catalysts, and membranes. Cold Spring Harb Symp Quant Biol 52: 805–24.

Cavalier-Smith, T. (2001) Obcells as Proto-Organisms: Membrane Heredity, Lithophosphorylation, and the Origins of the Genetic Code, the First Cells, and Photosynthesis. J Mol Evol 53: 555–595.

Cavalier-Smith, T. (2006) Rooting the tree of life by transition analyses. Biol Direct 1: 19.

Darwin, C., and Darwin, F. (1888) he life and letters of Charles Darwin, including an autobiographical chapter. .

Dwivedi, M.V., Harishchandra, R.K., Koshkina, O., Maskos, M., and Galla, H.-J. (2014) Size Influences the Effect of Hydrophobic Nanoparticles on Lung Surfactant Model Systems. Biophys J 106: 289–298.

Hentrich, C., and Szostak, J.W. (2014) Controlled Growth of Filamentous Fatty Acid Vesicles under Flow. Langmuir 30: 14916–14925.

Houston, P.L. (2006) Chemical kinetics and reaction dynamics. Dover Publications, .

Koonin, E. V. (2014) The origins of cellular life. Antonie Van Leeuwenhoek 106: 27–41.

Kurihara, K., Tamura, M., Shohda, K., Toyota, T., Suzuki, K., and Sugawara, T. (2011) Self-reproduction of supramolecular giant vesicles combined with the amplification of encapsulated DNA. Nat Chem 3: 775–781.

Mansy, S.S. (2010) Membrane Transport in Primitive Cells. Cold Spring Harb Perspect Biol 2: a002188–a002188.

Materese, C.K., Nuevo, M., and Sandford, S.A. (2017) The Formation of Nucleobases from the Ultraviolet Photoirradiation of Purine in Simple Astrophysical Ice Analogues. Astrobiology ast.2016.1613.

Rochette, É., Courtemanche, M.-A., Pulis, A., Bi, W., and Fontaine, F.-G. (2015) Ambiphilic Frustrated Lewis Pair Exhibiting High Robustness and Reversible Water Activation: Towards the Metal-Free Hydrogenation of Carbon Dioxide. Molecules 20: 11902–11914.

Ruiz-Mirazo, K., Briones, C., and la Escosura, A. de (2014) Prebiotic Systems Chemistry: New Perspectives for the Origins of Life. Chem Rev 114: 285–366.

Schrum, J.P., Zhu, T.F., and Szostak, J.W. (2010) The origins of cellular life. Cold Spring Harb Perspect Biol 2: a002212.

Wächtershäuser, G. (1988) Before enzymes and templates: theory of surface metabolism. Microbiol Rev 52: 452–84.

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N., et al. (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature 462: 1056–1060.

Yao, Q., and Yuan, C. (2013) A Synthetic Study of Chiral α-Hydroxy- H -Phosphinates Based on Proline Catalysis. Chem - A Eur J 19: 6080–6088.

Zhang, X. V., Martin, S.T., Friend, C.M., Schoonen, M.A.A., and Holland, H.D. (2004) Mineral-Assisted Pathways in Prebiotic Synthesis: Photoelectrochemical Reduction of Carbon(+IV) by Manganese Sulfide. J Am Chem Soc 126: 11247–11253.

Bibliography

Abby, S.S., Cury, J., Guglielmini, J., Néron, B., Touchon, M., and Rocha, E.P.C. (2016) Identification of protein secretion systems in bacterial genomes. *Sci Rep* **6**: 23080.

Abdeljabbar, H., Cayol, J.-L., Hania, W. Ben, Boudabous, A., Sadfi, N., Fardeau, M.-L., and Marie-Laure Fardeau, C. (2013) Halanaerobium sehlinense sp. nov., an extremely halophilic, fermentative, strictly anaerobic bacterium from sediments of the hypersaline lake Sehline Sebkha. *Int J Syst Evol Microbiol* 2069–7.

Adu-Bobie, J., Lupetti, P., Brunelli, B., Granoff, D., Norais, N., Ferrari, G., *et al.* (2004) GNA33 of Neisseria meningitidis is a lipoprotein required for cell separation, membrane architecture, and virulence. *Infect Immun* **72**: 1914–9.

Al-Hinai, M.A., Jones, S.W., and Papoutsakis, E.T. (2015) The Clostridium sporulation programs: diversity and preservation of endospore differentiation. *Microbiol Mol Biol Rev* **79**: 19–37.

Alauzet, C., Marchandin, H., Courtin, P., Mory, F., Lemée, L., Pons, J.-L., *et al.* (2014) Multilocus analysis reveals diversity in the genus Tissierella: Description of Tissierella carlieri sp. nov. in the new class Tissierellia classis nov. *Syst Appl Microbiol* **37**: 23–34.

Albrecht, R., Schütz, M., Oberhettinger, P., Faulstich, M., Bermejo, I., Rudel, T., *et al.* (2014) Structure of BamA, an essential factor in outer membrane protein biogenesis. *Acta Crystallogr Sect D Biol Crystallogr* **70**: 1779–1789.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–10.

Antunes, L.C., Poppleton, D., Klingl, A., Dupuy, B., Brochier-Armanetd, C., Beloin, C., and Gribaldo, S. (2016) Phylogenomic analysis supports the ancestral presence of LPS outer membranes in the Firmicutes. *Elife* e14589.

Arrieta, M.-C., Stiemsma, L.T., Dimitriu, P.A., Thorson, L., Russell, S., Yurist-Doutsch, S., *et al.* (2015) Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci Transl Med* **7**: 307ra152.

Aujoulat, F., Bouvet, P., Jumas-Bilak, E., Jean-Pierre, H., and Marchandin, H. (2014) Veillonella seminalis sp. nov., a novel anaerobic Gram-stain-negative coccus from human clinical samples, and emended description of the genus Veillonella. *Int J Syst Evol Microbiol* **64**: 3526–31.

Bagos, P.G., Liakopoulos, T.D., Spyropoulos, I.C., and Hamodrakas, S.J. (2004) PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. *Nucleic Acids Res* **32**: W400-4.

Bansal-Mutalik, R., and Nikaido, H. (2014) Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. *Proc Natl Acad Sci U S A* **111**: 4958–63.

Barreteau, H., Kovac, A., Boniface, A., Sova, M., Gobec, S., and Blanot, D. (2008) Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol Rev* **32**: 168–207.

Battistuzzi, F.U., and Hedges, S.B. (2009) A major clade of prokaryotes with ancient adaptations to life on land. *Mol Biol Evol* **26**: 335–43.

Bhattacharya, A., and Pletschke, B.I. (2014) Review of the enzymatic machinery of Halothermothrix orenii with special reference to industrial applications. *Enzyme Microb Technol* **55**: 159–169.

Bhatti, M. a, and Frank, M.O. (2000) Veillonella parvula meningitis: case report and review of Veillonella infections. *Clin Infect Dis* **31**: 839–840.

Bladen, H. a., and Mergenhagen, S.E. (1964) Ultrastructure of Veillonella and Morphological Correlation of an Outer Membrane with Particles Associated with Endotoxic Activity. *J Bacteriol* **88**: 1482–1492.

Blobel, G. (1980) Intracellular protein topogenesis. *Proc Natl Acad Sci U S A* **77**: 1496–500.

Boniface, A., Parquet, C., Arthur, M., Mengin-Lecreulx, D., and Blanot, D. (2009) The elucidation of the structure of Thermotoga maritima peptidoglycan reveals two novel types of cross-link. *J Biol Chem* **284**: 21856–62.

Bouhss, A., Trunkfield, A.E., Bugg, T.D.H., and Mengin-Lecreulx, D. (2008) The biosynthesis of peptidoglycan lipid-linked intermediates. *FEMS Microbiol Rev* **32**: 208–33.

Braun, V., and Rehn, K. (1969) Chemical characterization, spatial distribution and function of a lipoprotein (murein-lipoprotein) of the E. coli cell wall. The specific effect of trypsin on the membrane structure. *Eur J Biochem* **10**: 426–38.

Brown, S., Santa Maria, J.P., Walker, S., and Walker, S. (2013) Wall teichoic acids of gram-positive bacteria. *Annu Rev Microbiol* **67**: 313–36.

Buchanan, S.K. (2001) Type I secretion and multidrug

efflux: Transport through the TolC channel-tunnel. *Trends Biochem Sci* **26**: 3–6.

Buddelmeijer, N. (2015) The molecular mechanism of bacterial lipoprotein modification--how, when and why? *FEMS Microbiol Rev* **39**: 246–61.

Callaway, T.R., Dowd, S.E., Edrington, T.S., Anderson, R.C., Krueger, N., Bauer, N., *et al.* (2010) Evaluation of bacterial diversity in the rumen and feces of cattle fed different levels of dried distillers grains plus solubles using bacterial tag-encoded FLX amplicon pyrosequencing. *J Anim Sci* **88**: 3977–3983.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Campbell, C., Adeolu, M., and Gupta, R.S. (2015) Genome-based taxonomic framework for the class Negativicutes: division of the class Negativicutes into the orders Selenomonadales emend., Acidaminococcales ord. nov. and Veillonellales ord. nov. *Int J Syst Evol Microbiol* **65**: 3203–3215.

Campbell, C., Sutcliffe, I.C., and Gupta, R.S. (2014) Comparative proteome analysis of Acidaminococcus intestini supports a relationship between outer membrane biogenesis in Negativicutes and Proteobacteria. *Arch Microbiol* **196**: 307–10.

Campbell, J., Singh, A.K., Swoboda, J.G., Gilmore, M.S., Wilkinson, B.J., and Walker, S. (2012) An antibiotic that inhibits a late step in wall teichoic acid biosynthesis induces the cell wall stress stimulon in Staphylococcus aureus. *Antimicrob Agents Chemother* **56**: 1810–20.

Carlier, J.-P., Marchandin, H., Jumas-Bilak, E., Lorin, V., Henry, C., Carrière, C., and Jean-Pierre, H. (2002) Anaeroglobus geminatus gen. nov., sp. nov., a novel member of the family Veillonellaceae. *Int J Syst Evol Microbiol* **52**: 983–6.

Cascales, E., Bernadac, A., Gavioli, M., Lazzaroni, J.-C., and Lloubes, R. (2002) Pal lipoprotein of Escherichia coli plays a major role in outer membrane integrity. *J Bacteriol* **184**: 754–9.

Cava, F., and Pedro, M.A. de (2014) Peptidoglycan plasticity in bacteria: emerging variability of the murein sacculus and their associated biological functions. *Curr Opin Microbiol* **18**: 46–53.

Cavalier-Smith, T. (1987) The origin of cells: a symbiosis between genes, catalysts, and membranes. *Cold Spring Harb Symp Quant Biol* **52**: 805–24.

Cavalier-Smith, T. (2001) Obcells as Proto-Organisms: Membrane Heredity, Lithophosphorylation, and the Origins of the Genetic Code, the First Cells, and Photosynthesis. *J Mol Evol* **53**: 555–595.

Cavalier-Smith, T. (2002) The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int J Syst Evol Microbiol* **52**: 7–76.

Cavalier-Smith, T. (2006a) Rooting the tree of life by transition analyses. *Biol Direct* **1**: 19.

Cavalier-Smith, T. (2006b) Cell evolution and Earth history: stasis and revolution. *Philos Trans R Soc Lond B Biol Sci* **361**: 969–1006.

Cayol, J.L., Ollivier, B., Patel, B.K.C., Prensier, G., Guezennec, J., and Garcia, J.L. (1994) Isolation and Characterization of Halothemothrix orenii gen . *Int J Syst Bacteriol* **44**: 534–540.

Chen, Z.-Y., Cao, J., Xie, L., Li, X.-F., Yu, Z.-H., and Tong, W.-Y. (2014) Construction of leaky strains and extracellular production of exogenous proteins in recombinant Escherichia coli. *Microb Biotechnol* **7**: 360–70.

Chevrot, R., Carlotti, A., Sopena, V., Marchand, P., and Rosenfeld, E. (2008) Megamonas rupellensis sp. nov., an anaerobe isolated from the caecum of a duck. *Int J Syst Evol Microbiol* **58**: 2921–2924.

Ching, G., and Inouye, M. (1985) Evolution of the lipoprotein gene in the enterobacteriaceae. Cloning and DNA sequence of the lpp gene from Proteus mirabilis. *J Mol Biol* **185**: 501–7.

Collin, S., Guilvout, I., Nickerson, N.N., and Pugsley, A.P. (2011) Sorting of an integral outer membrane protein via the lipoprotein-specific Lol pathway and a dedicated lipoprotein pilotin. *Mol Microbiol* **80**: 655–665.

Confer, A.W., and Ayalew, S. (2013) The OmpA family of proteins: Roles in bacterial pathogenesis and immunity. *Vet Microbiol* **163**: 207–222.

Costerton, J.W., Damgaard, H.N., and Cheng, K.J. (1974) Cell envelope morphology of rumen bacteria. *J Bacteriol* **118**: 1132–43.

Darwin, C., and Darwin, F. (1888) *he life and letters of Charles Darwin, including an autobiographical chapter*. .

DebRoy, C., Fratamico, P.M., Yan, X., Baranzoni, G., Liu, Y., Needleman, D.S., *et al.* (2016) Comparison of O-Antigen Gene Clusters of All O-Serogroups of Escherichia coli and Proposal for Adopting a New Nomenclature for O-Typing. *PLoS One* **11**: e0147434.

Dong, Y., Sanford, R.A., Boyanov, M.I., Kemner, K.M., Flynn, T.M., O'Loughlin, E.J., *et al.* (2016) *Orenia*

metallireducens sp. nov. strain Z6, a Novel Metal-reducing *Firmicute* from the Deep Subsurface. *Appl Environ Microbiol* **82**: AEM.02382-16.

Downes, J., Munson, M., and Wade, W.G. (2003) Dialister invisus sp. nov., isolated from the human oral cavity. *Int J Syst Evol Microbiol* **53**: 1937–40.

Downes, J., Olsvik, B., Hiom, S.J., Spratt, D.A., Cheeseman, S.L., Olsen, I., *et al.* (2000) Bulleidia extructa gen. nov., sp. nov., isolated from the oral cavity. *Int J Syst Evol Microbiol* **50 Pt 3**: 979–83.

Dramsi, S., Magnet, S., Davison, S., and Arthur, M. (2008) Covalent attachment of proteins to peptidoglycan. *FEMS Microbiol Rev* **32**: 307–320.

Durai, P., Batool, M., and Choi, S. (2015) Structure and Effects of Cyanobacterial Lipopolysaccharides. *Mar Drugs* **13**: 4217–30.

Dwivedi, M.V., Harishchandra, R.K., Koshkina, O., Maskos, M., and Galla, H.-J. (2014) Size Influences the Effect of Hydrophobic Nanoparticles on Lung Surfactant Model Systems. *Biophys J* **106**: 289–298.

Eraso, J.M., Markillie, L.M., Mitchell, H.D., Taylor, R.C., Orr, G., and Margolin, W. (2014) The highly conserved MraZ protein is a transcriptional regulator in Escherichia coli. *J Bacteriol* **196**: 2053–66.

Errington, J., Mickiewicz, K., Kawai, Y., and Wu, L.J. (2016) L-form bacteria, chronic diseases and the origins of life. *Phil Trans R Soc B* **371**.

Euzeby, J. (1997) List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. *Int J Syst Bacteriol* 590–592 http://www.bacterio.net/veillonella.html. Accessed April 24, 2017.

Favini-Stabile, S., Contreras-Martel, C., Thielens, N., and Dessen, A. (2013) MreB and MurG as scaffolds for the cytoplasmic steps of peptidoglycan biosynthesis. *Environ Microbiol* **15**: 3218–3228.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., *et al.* (2014) Pfam: the protein families database. *Nucleic Acids Res* **42**: D222-30.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., *et al.* (2015) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**: D279–D285.

Fuerst, J.A. (2013) *Planctomycetes : cell structure, origins and biology*. Humana Press, .

Fukui, M., Kojima, H., and Watanabe, M. (2016) Complete genome sequence and cell structure of Limnochorda pilosa, a Gram-negative spore-former within the phylum Firmicutes. *Int J Syst Evol Microbiol* **66**: 1330–1339.

Fulghum, R.S., and Moore, W.E. (1963) Isolation, Enumeration, and Characteristics of proteoltic rumen bacteria. *J Bacteriol* **85**: 808–15.

Galán, J.C., Reig, M., Navas, A., Baquero, F., and Blázquez, J. (2000) ACI-1 from Acidaminococcus fermentans: characterization of the first beta-lactamase in Anaerobic cocci. *Antimicrob Agents Chemother* **44**: 3144–9.

Gautam, A., Vyas, R., and Tewari, R. (2011) Peptidoglycan biosynthesis machinery: a rich source of drug targets. *Crit Rev Biotechnol* **31**: 295–336.

Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balázsi, G., Ravasz, E., Daugherty, M.D., *et al.* (2003) Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J Bacteriol* **185**: 5673–84.

Godlewska, R., Wiśniewska, K., Pietras, Z., and Jagusztyn-Krynicka, E.K. (2009) Peptidoglycan-associated lipoprotein (Pal) of Gram-negative bacteria: function, structure, role in pathogenesis and potential application in immunoprophylaxis. *FEMS Microbiol Lett* **298**: 1–11.

Goldfine, H., and Shen, H. (eds) (2007) *Listeria monocytogenes: Pathogenesis and Host Response*. Springer US, Boston, MA.

Goolab, S., Roth, R.L., Heerden, H. van, and Crampton, M.C. (2015) Analyzing the molecular mechanism of lipoprotein localization in Brucella. *Front Microbiol* **6**: 1189.

Gram, C. (1884) Uber die isolirte Ffirbung der Schizomyceten in Schnitt- und Trockenpraparaten. *Fortshr Med* 185–189.

Green, E.R., and Mecsas, J. (2016) Bacterial Secretion Systems: An Overview. *Microbiol Spectr* **4**.

Griffen, A.L., Beall, C.J., Campbell, J.H., Firestone, N.D., Kumar, P.S., Yang, Z.K., *et al.* (2012) Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *ISME J* **6**: 1176–1185.

Grijpstra, J., Arenas, J., Rutten, L., and Tommassen, J. (2013) Autotransporter secretion: Varying on a theme. *Res Microbiol* **164**: 562–582.

Guo, M.S., Updegrove, T.B., Gogol, E.B., Shabalina, S.A., Gross, C.A., and Storz, G. (2014) MicL, a new σE-dependent sRNA, combats envelope stress by repressing

synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev* **28**: 1620–34.

Gupta, R.S. (2011) Origin of diderm (Gram-negative) bacteria: Antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes. *Antonie van Leeuwenhoek, Int J Gen Mol Microbiol* **100**: 171–182.

Gutierrez, J., Smith, R., and Pogliano, K. (2010) SpoIID-mediated peptidoglycan degradation is required throughout engulfment during Bacillus subtilis sporulation. *J Bacteriol* **192**: 3174–86.

Hamilton-Brehm, S.D., Mosher, J.J., Vishnivetskaya, T., Podar, M., Carroll, S., Allman, S., *et al.* (2010) Caldicellulosiruptor obsidiansis sp. nov., an anaerobic, extremely thermophilic, cellulolytic bacterium isolated from Obsidian Pool, Yellowstone National Park. *Appl Environ Microbiol* **76**: 1014–20.

Hammerschmidt, S., Wolff, S., Hocke, A., Rosseau, S., Müller, E., and Rohde, M. (2005) Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. *Infect Immun* **73**: 4653–67.

Hansel, A., Pattus, F., Jürgens, U.J., and Tadros, M.H. (1998) Cloning and characterization of the genes coding for two porins in the unicellular cyanobacterium Synechococcus PCC 6301. *Biochim Biophys Acta - Gene Struct Expr* **1399**: 31–39.

Hasegawa, K., Linnemann, R.W., Mansbach, J.M., Ajami, N.J., Espinola, J.A., Petrosino, J.F., *et al.* (2016) The Fecal Microbiota Profile and Bronchiolitis in Infants. *Pediatrics* **138**.

Hayat, S., Peters, C., Shu, N., Tsirigos, K.D., and Elofsson, A. (2016) Inclusion of dyad-repeat pattern improves topology prediction of transmembrane β-barrel proteins. *Bioinformatics* **32**: 1571–3.

Heinz, E., Selkrig, J., Belousoff, M., and Lithgow, T. (2015) Evolution of the Translocation and Assembly Module (TAM). *Genome Biol Evol* evv097.

Helander, I.M., Haikara, A., Sadovskaya, I., Vinogradov, E., and Salkinoja-Salonen, M.S. (2004) Lipopolysaccharides of anaerobic beer spoilage bacteria of the genus *Pectinatus* – lipopolysaccharides of a Gram-positive genus. *FEMS Microbiol Rev* **28**: 543–552.

Henderson, C., and Hodgkiss, W. (1973) An electron microscopic study of Anaerovibrio lipolytica (strain 5S) and its lipolytic enzyme. *J Gen Microbiol* **76**: 389–93.

Hentrich, C., and Szostak, J.W. (2014) Controlled Growth of Filamentous Fatty Acid Vesicles under Flow. *Langmuir* **30**: 14916–14925.

Higgins, M.L., and Shockman, G.D. (1970) Model for Cell Wall Growth of Streptococcus faecalis. *J Bacteriol* **101**: 643–648.

Hilbert, D.W., and Piggot, P.J. (2004) Compartmentalization of gene expression during Bacillus subtilis spore formation. *Microbiol Mol Biol Rev* **68**: 234–62.

Houston, P.L. (2006) *Chemical kinetics and reaction dynamics*. Dover Publications, .

Imachi, H., Sekiguchi, Y., Kamagata, Y., Hanada, S., Ohashi, A., and Harada, H. (2002) Pelotomaculum thermopropionicum gen. nov., sp. nov., an anaerobic, thermophilic, syntrophic propionate-oxidizing bacterium. *Int J Syst Evol Microbiol* **52**: 1729–35.

Iqbal, H., Kenedy, M.R., Lybecker, M., and Akins, D.R. (2016) The TamB ortholog of *Borrelia burgdorferi* interacts with the β-barrel assembly machine (BAM) complex protein BamA. *Mol Microbiol* **102**: 757–774.

Jacob-Dubuisson, F., Guérin, J., Baelen, S., and Clantin, B. (2013) Two-partner secretion: as simple as it sounds? *Res Microbiol* **164**: 583–595.

Jacquier, N., Viollier, P.H., and Greub, G. (2015) The role of peptidoglycan in chlamydial cell division: towards resolving the chlamydial anomaly. *FEMS Microbiol Rev* **39**: 262–275.

Jamet, S., Quentin, Y., Coudray, C., Texier, P., Laval, F., Daffé, M., *et al.* (2015) Evolution of Mycolic Acid Biosynthesis Genes and Their Regulation during Starvation in Mycobacterium tuberculosis. *J Bacteriol* **197**: 3797–811.

Jeeves, M., Sridhar, P., and Knowles, T.J. (2015) Expression, Purification, and Screening of BamE, a Component of the BAM Complex, for Structural Characterization. In *Methods in molecular biology (Clifton, N.J.)*. pp. 245–258.

Joseleau-Petit, D., Liébart, J.-C., Ayala, J.A., and D'Ari, R. (2007) Unstable Escherichia coli L forms revisited: growth requires peptidoglycan synthesis. *J Bacteriol* **189**: 6512–20.

Jumas-Bilak, E., Carlier, J.-P., Jean-Pierre, H., Mory, F., Teyssier, C., Gay, B., *et al.* (2007) Acidaminococcus intestini sp. nov., isolated from human clinical samples. *Int J Syst Evol Microbiol* **57**: 2314–9.

Jumas-Bilak, E., Jean-Pierre, H., Carlier, J.-P., Teyssier, C., Bernard, K., Gay, B., *et al.* (2005) Dialister micraerophilus sp. nov. and Dialister propionicifaciens sp. nov., isolated from human clinical samples. *Int J Syst Evol Microbiol* **55**: 2471–8.

Jumas-Bilak, E., Roudiere, L., and Marchandin, H.

(2009) Description of “Synergistetes” phyl. nov. and emended description of the phylum “Deferribacteres” and of the family Syntrophomonadaceae, phylum “Firmicutes.” *Int J Syst Evol Microbiol* **59**: 1028–1035.

Jürgens, U.J., Drews, G., and Weckesser, J. (1983) Primary structure of the peptidoglycan from the unicellular cyanobacterium Synechocystis sp. strain PCC 6714. *J Bacteriol* **154**: 471–8.

Kalmokoff, M.L., Austin, J.W., Cyr, T.D., Hefford, M.A., Teather, R.M., and Selinger, L.B. (2009) Physical and genetic characterization of an outer-membrane protein (OmpM1) containing an N-terminal S-layer-like homology domain from the phylogenetically Gram-positive gut anaerobe Mitsuokella multacida. *Anaerobe* **15**: 74–81.

Kamio, Y., and Nakamura, K. (1987) Putrescine and cadaverine are constituents of peptidoglycan in Veillonella alcalescens and Veillonella parvula. *J Bacteriol* **169**: 2881–4.

Kane, M.D., and Breznak, J.A. (1991) Acetonema longum gen. nov. sp. nov., an H2/CO2 acetogenic bacterium from the termite, Pterotermes occidentis. *Arch Microbiol* **156**: 91–8.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**: D353–D361.

Kim, Y.C., Tarr, A.W., and Penfold, C.N. (2014) Colicin import into E. coli cells: A model system for insights into the import mechanisms of bacteriocins. *Biochim Biophys Acta - Mol Cell Res* **1843**: 1717–1731.

Knapp, S., Brodal, C., Peterson, J., Qi, F., Kreth, J., and Merritt, J. (2017) Natural Competence Is Common among Clinical Isolates of Veillonella parvula and Is Useful for Genetic Manipulation of This Key Member of the Oral Microbiome. *Front Cell Infect Microbiol* **7**: 139.

Knowles, T.J., Scott-Tucker, A., Overduin, M., and Henderson, I.R. (2009) Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nat Rev Microbiol* **7**: 206–14.

Kodani, S., Ishida, K., and Murakami, M. (1999) Occurrence and identification of UDP-N-acetylmuramyl-pentapeptide from the cyanobacterium Anabaena cylindrica. *FEMS Microbiol Lett* **176**.

Koizumi, N., and Watanabe, H. (2003) Molecular cloning and characterization of a novel leptospiral lipoprotein with OmpA domain. *FEMS Microbiol Lett* **226**: 215–9.

Kojima, S., and Kamio, Y. (2012) Molecular basis for the maintenance of envelope integrity in Selenomonas ruminantium: cadaverine biosynthesis and covalent modification into the peptidoglycan play a major role. *J Nutr Sci Vitaminol (Tokyo)* **58**: 153–60.

Kojima, S., Kaneko, J., Abe, N., Takatsuka, Y., and Kamio, Y. (2011) Cadaverine covalently linked to the peptidoglycan serves as the correct constituent for the anchoring mechanism between the outer membrane and peptidoglycan in Selenomonas ruminantium. *J Bacteriol* **193**: 2347–2350.

Koonin, E. V. (2014) The origins of cellular life. *Antonie Van Leeuwenhoek* **106**: 27–41.

Krewulak, K.D., and Vogel, H.J. (2011) TonB or not TonB: is that the question? *Biochem Cell Biol* **89**: 87–97.

Kurihara, K., Tamura, M., Shohda, K., Toyota, T., Suzuki, K., and Sugawara, T. (2011) Self-reproduction of supramolecular giant vesicles combined with the amplification of encapsulated DNA. *Nat Chem* **3**: 775–781.

Lake, J. a (2009) Evidence for an early prokaryotic endosymbiosis. *Nature* **460**: 967–971.

Layec, S., Decaris, B., and Leblond-Bourget, N. (2008) Diversity of Firmicutes peptidoglycan hydrolases and specificities of those involved in daughter cell separation. *Res Microbiol* **159**: 507–15.

Leaver, M., Domínguez-Cuevas, P., Coxhead, J.M., Daniel, R.A., and Errington, J. (2009) Life without a wall or division machine in Bacillus subtilis. *Nature* **457**: 849–53.

Lee, J., Kim, D.H., and Hwang, I. (2014) Specific targeting of proteins to outer envelope membranes of endosymbiotic organelles, chloroplasts, and mitochondria. *Front Plant Sci* **5**: 173.

Lee, Y.J., Romanek, C.S., Mills, G.L., Davis, R.C., Whitman, W.B., and Wiegel, J. (2006) Gracilibacter thermotolerans gen. nov., sp. nov., an anaerobic, thermotolerant bacterium from a constructed wetland receiving acid sulfate water. *Int J Syst Evol Microbiol* **56**: 2089–2093.

Leewenhoeck, A. (1677) Observation, communicated to the publisher by Mr. Antony van Leewenhoeck, in a Dutch letter of the 9 Octob. 1676 here English’d: concerning little animals by him observed in rain-well-sea and snow water; as also in water wherein pepper had lain infused. *Philos Trans R Soc London B Biol Sci* **12**: 821–831.

León, K.B. De, Utturkar, S.M., Camilleri, L.B., Elias, D.A., Arkin, A.P., Fields, M.W., *et al.* (2015) Complete Genome Sequence of Pelosinus fermentans JBW45, a Member of a Remarkably Competitive Group of Negativicutes in the Firmicutes Phylum. *Genome Announc* **3**.

Leyton, D.L., Rossiter, A.E., and Henderson, I.R. (2012) From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis. *Nat Rev Microbiol* **10**: 213–225.

Liang, R., Davidova, I.A., Marks, C.R., Stamps, B.W., Harriman, B.H., Stevenson, B.S., *et al.* (2016) Metabolic Capability of a Predominant Halanaerobium sp. in Hydraulically Fractured Gas Wells and Its Implication in Pipeline Corrosion. *Front Microbiol* **7**: 988.

Limam, R.D., Bouchez, T., Chouari, R., Li, T., Barkallah, I., Landoulsi, A., and Sghir, A. (2010) Detection of WWE2-related Lentisphaerae by 16S rRNA gene sequencing and fluorescence in situ hybridization in landfill leachate. *Can J Microbiol* **56**: 846–52.

Liu, J., Xie, Z., Merritt, J., and Qi, F. (2012) Establishment of a tractable genetic transformation system in Veillonella spp. *Appl Environ Microbiol* **78**: 3488–91.

Llamas, M.A., Ramos, J.L., and Rodríguez-Herva, J.J. (2000) Mutations in each of the tol genes of Pseudomonas putida reveal that they are critical for maintenance of outer membrane stability. *J Bacteriol* **182**: 4764–72.

Lorenzo, F. Di, Palmigiano, A., Paciello, I., Pallach, M., Garozzo, D., Bernardini, M.-L., *et al.* (2017) The Deep-Sea Polyextremophile Halobacteroides lacunaris TB21 Rough-Type LPS: Structure and Inhibitory Activity towards Toxic LPS. *Mar Drugs* **15**: 201.

Lupas, A., Müller, S., Goldie, K., Engel, A.M., Engel, A., and Baumeister, W. (1995) Model Structure of the Ompα Rod, a Parallel Four-stranded Coiled Coil from the Hyperthermophilic EubacteriumThermotoga maritima. *J Mol Biol* **248**: 180–189.

Machowski, E.E., Senzani, S., Ealand, C., and Kana, B.D. (2014) Comparative genomics for mycobacterial peptidoglycan remodelling enzymes reveals extensive genetic multiplicity. *BMC Microbiol* **14**: 75.

Magnet, S., Dubost, L., Marie, A., Arthur, M., and Gutmann, L. (2008) Identification of the L,D-transpeptidases for peptidoglycan cross-linking in Escherichia coli. *J Bacteriol* **190**: 4782–5.

Males, B.M., Berthold, P., Dougherty, P.A., and Listgarten, M.A. (1984) Helical flagellation in Centipeda periodontii, a gram-negative, anaerobic bacillus from periodontitis lesions. *J Gen Microbiol* **130**: 185–91.

Mansy, S.S. (2010) Membrane Transport in Primitive Cells. *Cold Spring Harb Perspect Biol* **2**: a002188–a002188.

Marchandin, H., Jumas-Bilak, E., Gay, B., Teyssier, C., Jean-Pierre, H., Buochberg, M.S. de, *et al.* (2003) Phylogenetic analysis of some Sporomusa sub-branch members isolated from human clinical specimens: description of Megasphaera micronuciformis sp. nov. *Int J Syst Evol Microbiol* **53**: 547–53.

Marchandin, H., Teyssier, C., Campos, J., Jean-Pierre, H., Roger, F., Gay, B., *et al.* (2010) Negativicoccus succinicivorans gen. nov., sp. nov., isolated from human clinical samples, emended description of the family Veillonellaceae and description of Negativicutes classis nov., Selenomonadales ord. nov. and Acidaminococcaceae fam. nov. in the ba. *Int J Syst Evol Microbiol* **60**: 1271–1279.

Matera, G., Muto, V., Vinci, M., Zicca, E., Abdollahi-Roodsaz, S., Veerdonk, F.L. van de, *et al.* (2009) Receptor recognition of and immune intracellular pathways for Veillonella parvula lipopolysaccharide. *Clin Vaccine Immunol* **16**: 1804–9.

Materese, C.K., Nuevo, M., and Sandford, S.A. (2017) The Formation of Nucleobases from the Ultraviolet Photoirradiation of Purine in Simple Astrophysical Ice Analogues. *Astrobiology* ast.2016.1613.

Matias, V.R.F., and Beveridge, T.J. (2006) Native cell wall organization shown by cryo-electron microscopy confirms the existence of a periplasmic space in Staphylococcus aureus. *J Bacteriol* **188**: 1011–21.

Matias, V.R.F., and Beveridge, T.J. (2008) Lipoteichoic acid is a major component of the Bacillus subtilis periplasm. *J Bacteriol* **190**: 7414–8.

Mavromatis, K., Ivanova, N., Anderson, I., Lykidis, A., Hooper, S.D., Sun, H., *et al.* (2009) Genome analysis of the anaerobic thermohalophilic bacterium Halothermothrix orenii. *PLoS One* **4**.

Mayer, F., Lurz, R., and Schoberth, S. (1977) Electron microscopic investigation of the hydrogen-oxidizing acetate-forming anaerobic bacterium Acetobacterium woodii. *Arch Microbiol* **115**: 207–13.

Mesbah, N.M., Hedrick, D.B., Peacock, A.D., Rohde, M., and Wiegel, J. (2007) Natranaerobius thermophilus gen. nov., sp. nov., a halophilic, alkalithermophilic bacterium from soda lakes of the Wadi An Natrun, Egypt, and proposal of Natranaerobiaceae fam. nov. and Natranaerobiales ord. nov. *Int J Syst Evol Microbiol* **57**: 2507–12.

Misra, C.S., Basu, B., and Apte, S.K. (2015) Surface (S)-layer proteins of Deinococcus radiodurans and their utility as vehicles for surface localization of functional proteins. *Biochim Biophys Acta - Biomembr* **1848**: 3181–3187.

Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., and Punta, M. (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* **41**: e121.

Mohammadi, T., Dam, V. van, Sijbrandi, R., Vernet, T., Zapun, A., Bouhss, A., *et al.* (2011) Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. *EMBO J* **30**: 1425–32.

Mohammadi, T., Karczmarek, A., Crouvoisier, M., Bouhss, A., Mengin-Lecreulx, D., and Blaauwen, T. den (2007) The essential peptidoglycan glycosyltransferase MurG forms a complex with proteins involved in lateral envelope growth as well as with proteins involved in cell division in Escherichia coli. *Mol Microbiol* **65**: 1106–21.

Mori, K., Kaoru, Y., Yayoi, S., Urabe, T., and Suzuki, K. (2009) Caldisericum exile gen. nov., sp. nov., an anaerobic, thermophilic, filamentous bacterium of a novel bacterial phylum, Caldiserica phyl. nov., originally called the candidate phylum OP5, and description of Caldisericaceae fam. nov., Caldisericales ord. no. *Int J Syst Evol Microbiol* **59**: 2894–8.

Morotomi, M., Nagai, F., Sakon, H., and Tanaka, R. (2008) Dialister succinatiphilus sp. nov. and Barnesiella intestinihominis sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* **58**: 2716–20.

Morrison, M., and Miron, J. (2000) Adhesion to cellulose by Ruminococcus albus : a combination of cellulosomes and Pil-proteins? *FEMS Microbiol Lett* **185**: 109–115.

Mukherjee, S., and Kearns, D.B. (2014) The Structure and Regulation of Flagella in Bacillus subtilis. *Annu Rev Genet* **48**: 319–340.

Muñoz-Gómez, S.A., and Roger, A.J. (2016) Leaving negative ancestors behind. *Elife* **5**.

Nakayama, H., Kurokawa, K., and Lee, B.L. (2012) Lipoproteins in bacteria: Structures and biosynthetic pathways. *FEBS J* **279**: 4247–4268.

Needham, B.D., and Trent, M.S. (2013) Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. *Nat Rev Microbiol* **11**: 467–481.

Nivaskumar, M., and Francetic, O. (2014) Type II secretion system: A magic beanstalk or a protein escalator. *Biochim Biophys Acta - Mol Cell Res* **1843**: 1568–1577.

Noinaj, N., Fairman, J.W., and Buchanan, S.K. (2011) The Crystal Structure of BamB Suggests Interactions with BamA and Its Role within the BAM Complex. *J Mol Biol* **407**: 248–260.

Noinaj, N., Gumbart, J.C., and Buchanan, S.K. (2017) The β-barrel assembly machinery in motion. *Nat Rev Microbiol* **15**: 197–204.

Oberto, J. (2013) SyntTax: a web server linking synteny to prokaryotic taxonomy. *BMC Bioinformatics* **14**: 4.

Opiyo, S.O., Pardy, R.L., Moriyama, H., and Moriyama, E.N. (2010) Evolution of the Kdo2-lipid A biosynthesis in bacteria. *BMC Evol Biol* **10**: 362.

Park, J.S., Lee, W.C., Yeo, K.J., Ryu, K.-S., Kumarasiri, M., Hesek, D., *et al.* (2012) Mechanism of anchoring of OmpA protein to the cell wall peptidoglycan of the gram-negative bacterial outer membrane. *FASEB J* **26**: 219–28.

Pei, J., and Grishin, N. V (2005) COG3926 and COG5526: a tale of two new lysozyme-like protein families. *Protein Sci* **14**: 2574–81.

Pelicic, V. (2008) Type IV pili: e pluribus unum? *Mol Microbiol* **68**: 827–837.

Periasamy, S., and Kolenbrander, P.E. (2010) Central role of the early colonizer Veillonella sp. in establishing multispecies biofilm communities with initial, middle, and late colonizers of enamel. *J Bacteriol* **192**: 2965–2972.

Permpoonpattana, P., Phetcharaburanin, J., Mikelsone, A., Dembek, M., Tan, S., Brisson, M.-C., *et al.* (2013) Functional characterization of Clostridium difficile spore coat proteins. *J Bacteriol* **195**: 1492–503.

Petrus, A.K., Swithers, K.S., Ranjit, C., Wu, S., Brewer, H.M., Gogarten, J.P., *et al.* (2012) Genes for the major structural components of Thermotogales species' togas revealed by proteomic and evolutionary analyses of OmpA and OmpB homologs. *PLoS One* **7**: e40236.

Pinho, M.G., Kjos, M., and Veening, J.-W. (2013) How to get (a)round: mechanisms controlling growth and division of coccoid bacteria. *Nat Rev Microbiol* **11**: 601–14.

Polissi, A., and Sperandeo, P. (2014) The lipopolysaccharide export pathway in Escherichia coli: Structure, organization and regulated assembly of the Lpt machinery. *Mar Drugs* **12**: 1023–1042.

Pomeroy, C., Shanholtzer, C.J., and Peterson, L.R. (1987) Selenomonas bacteraemia--case report and review of the literature. *J Infect* **15**: 237–42.

Poppleton, D.I., Duchateau, M., Hourdel, V., Matondo, M., Flechsler, J., Klingl, A., *et al.* (2017) Outer Membrane Proteome of Veillonella parvula: A Diderm Firmicute of the Human Microbiome. *Front Microbiol* **8**: 1215.

Prat, L., Heinemann, I.U., Aerni, H.R., Rinehart, J., O'Donoghue, P., and Soll, D. (2012) Carbon source-dependent expansion of the genetic code in bacteria. *Proc Natl Acad Sci* **109**: 21070–21075.

Quintela, J.C., García-del Portillo, F., Pittenauer, E., Allmaier, G., and Pedro, M.A. de (1999) Peptidoglycan fine structure of the radiotolerant bacterium Deinococcus

radiodurans Sark. *J Bacteriol* **181**: 334–7.

Rauschenbach, I., Narasingarao, P., and Haggblom, M.M. (2011) Desulfurispirillum indicum sp. nov., a selenate- and selenite-respiring bacterium isolated from an estuarine canal. *Int J Syst Evol Microbiol* **61**: 654–658.

Regulski, K., Courtin, P., Meyrand, M., Claes, I.J.J., Lebeer, S., Vanderleyden, J., *et al.* (2012) Analysis of the peptidoglycan hydrolase complement of Lactobacillus casei and characterization of the major γ-D-glutamyl-L-lysyl-endopeptidase. *PLoS One* **7**: e32301.

Rochette, É., Courtemanche, M.-A., Pulis, A., Bi, W., and Fontaine, F.-G. (2015) Ambiphilic Frustrated Lewis Pair Exhibiting High Robustness and Reversible Water Activation: Towards the Metal-Free Hydrogenation of Carbon Dioxide. *Molecules* **20**: 11902–11914.

Rogosa, M. (1969) Acidaminococcus gen. n., Acidaminococcus fermentans sp. n., anaerobic gram-negative diplococci using amino acids as the sole energy source for growth. *J Bacteriol* **98**: 756–66.

Rossiter, A.E., Leyton, D.L., Tveen-Jensen, K., Browning, D.F., Sevastsyanovich, Y., Knowles, T.J., *et al.* (2011) The essential β-barrel assembly machinery complex components BamD and BamA are required for autotransporter biogenesis. *J Bacteriol* **193**: 4250–3.

Roush, D., Elias, D., and Mormile, M. (2014) Metabolic Capabilities of the Members of the Order Halanaerobiales and Their Potential Biotechnological Applications. *Curr Biotechnol* **3**: 3–9.

Ruiz-Mirazo, K., Briones, C., and la Escosura, A. de (2014) Prebiotic Systems Chemistry: New Perspectives for the Origins of Life. *Chem Rev* **114**: 285–366.

Russel, M. (1994) Mutants at conserved positions in gene IV, a gene required for assembly and secretion of filamentous phages. *Mol Microbiol* **14**: 357–69.

Sauvage, E., Kerff, F., Terrak, M., Ayala, J.A., and Charlier, P. (2008a) The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol Rev* **32**: 234–58.

Sauvage, E., Kerff, F., Terrak, M., Ayala, J.A., and Charlier, P. (2008b) The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol Rev* **32**: 234–58.

Schleifer, K.H., and Kandler, O. (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev* **36**: 407–77.

Schneewind, O., and Missiakas, D. (2014) Lipoteichoic acids, phosphate-containing polymers in the envelope of gram-positive bacteria. *J Bacteriol* **196**: 1133–1142.

Schneewind, O., and Missiakas, D.M. (2012) Protein secretion and surface display in Gram-positive bacteria. *Philos Trans R Soc Lond B Biol Sci* **367**: 1123–39.

Schoonmaker, M.K., Bishai, W.R., and Lamichhane, G. (2014) Nonclassical transpeptidases of Mycobacterium tuberculosis alter cell size, morphology, the cytosolic matrix, protein localization, virulence, and resistance to β-lactams. *J Bacteriol* **196**: 1394–402.

Schrum, J.P., Zhu, T.F., and Szostak, J.W. (2010) The origins of cellular life. *Cold Spring Harb Perspect Biol* **2**: a002212.

Schwechheimer, C., Kulp, A., Kuehn, M.J., Berleman, J., Auer, M., Deatherage, B., *et al.* (2014) Modulation of bacterial outer membrane vesicle production by envelope structure and content. *BMC Microbiol* **14**: 324.

Selkrig, J., Leyton, D.L., Webb, C.T., and Lithgow, T. (2014) Assembly of β-barrel proteins into bacterial outer membranes. *Biochim Biophys Acta - Mol Cell Res* **1843**: 1542–1550.

Selkrig, J., Mosbahi, K., Webb, C.T., Belousoff, M.J., Perry, a. J., Wells, T.J., *et al.* (2012) Discovery of an archetypal protein transport system in bacterial outer membranes. *Nat Struct Mol Biol* **19**: 506–510.

Sha, J., Fadl, A.A., Klimpel, G.R., Niesel, D.W., Popov, V.L., and Chopra, A.K. (2004) The two murein lipoproteins of Salmonella enterica serovar Typhimurium contribute to the virulence of the organism. *Infect Immun* **72**: 3987–4003.

Sham, L.-T., Butler, E.K., Lebar, M.D., Kahne, D., Bernhardt, T.G., and Ruiz, N. (2014) Bacterial cell wall. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. *Science* **345**: 220–2.

Shi, F., Harada, T., Ogawa, Y., Ono, H., Ohnishi-Kameyama, M., Miyamoto, T., *et al.* (2012) Capsular polysaccharide of Erysipelothrix rhusiopathiae, the causative agent of swine erysipelas, and its modification with phosphorylcholine. *Infect Immun* **80**: 3993–4003.

Shu, W., Liu, J., Ji, H., and Lu, M. (2000) Core structure of the outer membrane lipoprotein from Escherichia coli at 1.9 A resolution. *J Mol Biol* **299**: 1101–12.

Silhavy, T.J., Kahne, D., and Walker, S. (2010) The Bacterial Cell Envelope. **2**: 16.

Simmerman, R.F., Dave, A.M., and Bruce, B.D. (2014) Structure and Function of POTRA Domains of Omp85/TPS Superfamily. In *International review of cell and molecular biology*. pp. 1–34.

Sokolova, T.G., González, J.M., Kostrikina, N.A., Chernyh, N.A., Slepova, T. V, Bonch-Osmolovskaya, E.A., and Robb, F.T. (2004) Thermosinus

carboxydivorans gen. nov., sp. nov., a new anaerobic, thermophilic, carbon-monoxide-oxidizing, hydrogenogenic bacterium from a hot pool of Yellowstone National Park. *Int J Syst Evol Microbiol* **54**: 2353–9.

Spirig, T., Weiner, E.M., and Clubb, R.T. (2011) Sortase enzymes in Gram-positive bacteria. *Mol Microbiol* **82**: 1044–1059.

Strati, F., Cavalieri, D., Albanese, D., Felice, C. De, Donati, C., Hayek, J., *et al.* (2017) New evidences on the altered gut microbiota in autism spectrum disorders. *Microbiome* **5**: 24.

Sturgis, J.N. (2001) Organisation and evolution of the tol-pal gene cluster. *J Mol Microbiol Biotechnol* **3**: 113–22.

Sugawara, E., Nestorovich, E.M., Bezrukov, S.M., and Nikaido, H. (2006) Pseudomonas aeruginosa porin OprF exists in two different conformations. *J Biol Chem* **281**: 16220–9.

Sutcliffe, I.C. (2010) A phylum level perspective on bacterial cell envelope architecture. *Trends Microbiol* **18**: 464–470.

Sutcliffe, I.C. (2011) Cell envelope architecture in the Chloroflexi: a shifting frontline in a phylogenetic turf war. *Environ Microbiol* **13**: 279–282.

Sutcliffe, I.C., Harrington, D.J., and Hutchings, M.I. (2012) A phylum level analysis reveals lipoprotein biosynthesis to be a fundamental property of bacteria. *Protein Cell* **3**: 163–70.

Sveen, K., Hofstad, T., and Milner, K.C. (1977) Lethality for mice and chick embryos, pyrogenicity in rabbits and ability to gelate lysate from amoebocytes of Limulus polyphemus by lipopolysaccharides from Bacteroides, Fusobacterium and Veillonella. *Acta Pathol Microbiol Scand B* **85B**: 388–396.

Swithers, K.S., Fournier, G.P., Green, A.G., Gogarten, J.P., and Lapierre, P. (2011) Reassessment of the Lineage Fusion Hypothesis for the Origin of Double Membrane Bacteria. *PLoS One* **6**: e23774.

Takagi, A., Nakamura, K., and Ueda, M. (1965) Electron microscope studies of the intracytoplasmic membrane system in Clostridium tetani and Clostridium botulinum. *Jpn J Microbiol* **9**: 131–43.

Takeshita, T., Matsuo, K., Furuta, M., Shibata, Y., Fukami, K., Shimazaki, Y., *et al.* (2014) Distinct composition of the oral indigenous microbiota in South Korean and Japanese adults. *Sci Rep* **4**: 6990.

Tamaki, H., Tanaka, Y., Matsuzawa, H., Muramatsu, M., Meng, X.-Y., Hanada, S., *et al.* (2011) Armatimonas rosea gen. nov., sp. nov., of a novel bacterial phylum, Armatimonadetes phyl. nov., formally called the candidate phylum OP10. *Int J Syst Evol Microbiol* **61**: 1442–1447.

Tanner, A., Bouldin, H.D., and Maiden, M.F. (1989) Newly delineated periodontal pathogens with special reference to selenomonas species. *Infection* **17**: 182–7.

Tocheva, E.I., Matson, E.G., Morris, D.M., Moussavi, F., Leadbetter, J.R., and Jensen, G.J. (2011) Peptidoglycan remodeling and conversion of an inner membrane into an outer membrane during sporulation. *Cell* **146**: 799–812.

Tocheva, E.I., Ortega, D.R., and Jensen, G.J. (2016) Sporulation, bacterial cell envelopes and the origin of life. *Nat Rev Microbiol* **14**: 535–542.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., and Gordon, J.I. (2007) The human microbiome project. *Nature* **449**: 804–10.

Turner, R.D., Vollmer, W., and Foster, S.J. (2014) Different walls for rods and balls: the diversity of peptidoglycan. *Mol Microbiol* **91**: 862–74.

Typas, A., Banzhaf, M., Berg van Saparoea, B. van den, Verheul, J., Biboy, J., Nichols, R.J., *et al.* (2010) Regulation of Peptidoglycan Synthesis by Outer-Membrane Proteins. *Cell* **143**: 1097–1109.

Ulsen, P. van, Rahman, S. ur, Jong, W.S.P., Daleke-Schermerhorn, M.H., and Luirink, J. (2014) Type V secretion: From biogenesis to biotechnology. *Biochim Biophys Acta - Mol Cell Res* **1843**: 1592–1611.

Veillon, A., and Zuber, A. (1898) Sur Quelques Microbes Strictment Anaerobies et leur Role en Pathologie. *Arch Med Exp* 517–545.

Vollmer, W., Blanot, D., and Pedro, M.A. de (2008) Peptidoglycan structure and architecture. *FEMS Microbiol Rev* **32**: 149–67.

Volokhina, E.B., Grijpstra, J., Stork, M., Schilders, I., Tommassen, J., and Bos, M.P. (2011) Role of the periplasmic chaperones Skp, SurA, and DegQ in outer membrane protein biogenesis in Neisseria meningitidis. *J Bacteriol* **193**: 1612–1621.

Voulhoux, R., Bos, M.P., Geurtsen, J., Mols, M., and Tommassen, J. (2003) Role of a Highly Conserved Bacterial Protein in Outer Membrane Protein Assembly. *Science (80-)* **299**: 262–265.

Wächtershäuser, G. (1988) Before enzymes and templates: theory of surface metabolism. *Microbiol Rev* **52**: 452–84.

Wang, X., and Quinn, P.J. (2010) Lipopolysaccharide: Biosynthetic pathway and structure modification. *Prog*

Lipid Res **49**: 97–107.

Warner, L.R., Varga, K., Lange, O.F., Baker, S.L., Baker, D., Sousa, M.C., and Pardi, A. (2011) Structure of the BamC Two-Domain Protein Obtained by Rosetta with a Limited NMR Data Set. *J Mol Biol* **411**: 83–95.

Watanabe, M., Kojima, H., and Fukui, M. (2015) Limnochorda pilosa gen. nov., sp. nov., a moderately thermophilic, facultatively anaerobic, pleomorphic bacterium and proposal of Limnochordaceae fam. nov., Limnochordales ord. nov. and Limnochordia classis nov. in the phylum Firmicutes. *Int J Syst Evol Microbiol* **65**: 2378–2384.

Webb, C.T., Heinz, E., and Lithgow, T. (2012) Evolution of the ß-barrel assembly machinery. *Trends Microbiol* **20**: 612–620.

Wessel, A.K., Liew, J., Kwon, T., Marcotte, E.M., and Whiteley, M. (2013) Role of Pseudomonas aeruginosa peptidoglycan-associated outer membrane proteins in vesicle formation. *J Bacteriol* **195**: 213–9.

Whitman, W.B. (2011) *Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*. 2nd Editio., Springer, .

Wilson, E.O. (2013) *Letters to a young scientist*. .

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N., *et al.* (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462**: 1056–1060.

Wyatt, H.D.M., and West, S.C. (2014) Holliday junction resolvases. *Cold Spring Harb Perspect Biol* **6**: a023192.

Xue, Y., Xu, Y., Liu, Y., Ma, Y., and Zhou, P. (2001) Thermoanaerobacter tengcongensis sp. nov., a novel anaerobic, saccharolytic, thermophilic bacterium isolated from a hot spring in Tengcong, China. *Int J Syst Evol Microbiol* **51**: 1335–41.

Yamada, K.D., Tomii, K., and Katoh, K. (2016) Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees. *Bioinformatics* btw412.

Yang, C., Rooke, J.A., Cabeza, I., and Wallace, R.J. (2016) Nitrate and Inhibition of Ruminal Methanogenesis: Microbial Ecology, Obstacles, and Opportunities for Lowering Methane Emissions from Ruminant Livestock. *Front Microbiol* **7**: 132.

Yao, Q., and Yuan, C. (2013) A Synthetic Study of Chiral α-Hydroxy- *H* -Phosphinates Based on Proline Catalysis. *Chem - A Eur J* **19**: 6080–6088.

Zeikus, J.G., Hegge, P.W., Thompson, T.E., Phelps, T.J., and Langworthy, T.A. (1983) Isolation and description ofHaloanaerobium praevalens gen. nov. and sp. nov., an obligately anaerobic halophile common to Great Salt Lake sediments. *Curr Microbiol* **9**: 225–233.

Zeikus, J.G., Hegge, P.W., Thompson, T.E., Phelps, T.J., and Langworthy, T. a. (1983) Isolation and description of Haloanaerobium praevalens gen. nov. and sp. nov., an obligately anaerobic halophile common to Great Salt Lake sediments. *Curr Microbiol* **9**: 225–233.

Zhang, K., and Dong, X. (2009) Selenomonas bovis sp. nov., isolated from yak rumen contents. *Int J Syst Evol Microbiol* **59**: 2080–3.

Zhang, X. V., Martin, S.T., Friend, C.M., Schoonen, M.A.A., and Holland, H.D. (2004) Mineral-Assisted Pathways in Prebiotic Synthesis: Photoelectrochemical Reduction of Carbon(+IV) by Manganese Sulfide. *J Am Chem Soc* **126**: 11247–11253.

Zhou, P., Xie, G., Li, X., Liu, J., and Qi, F. (2017) Complete Genome Sequence of Veillonella atypica OK5, the First Transformable Strain in the Species. *Genome Announc* **5**: e00391-17.

Zückert, W.R. (2014) Secretion of Bacterial Lipoproteins: Through the Cytoplasmic Membrane, the Periplasm and Beyond. *Biochim Biophys Acta - Mol Cell Res* **1843**: 1509–1516.

Appendix – Figures

1.1 Supplemental Figures from Antunes et al. 2016

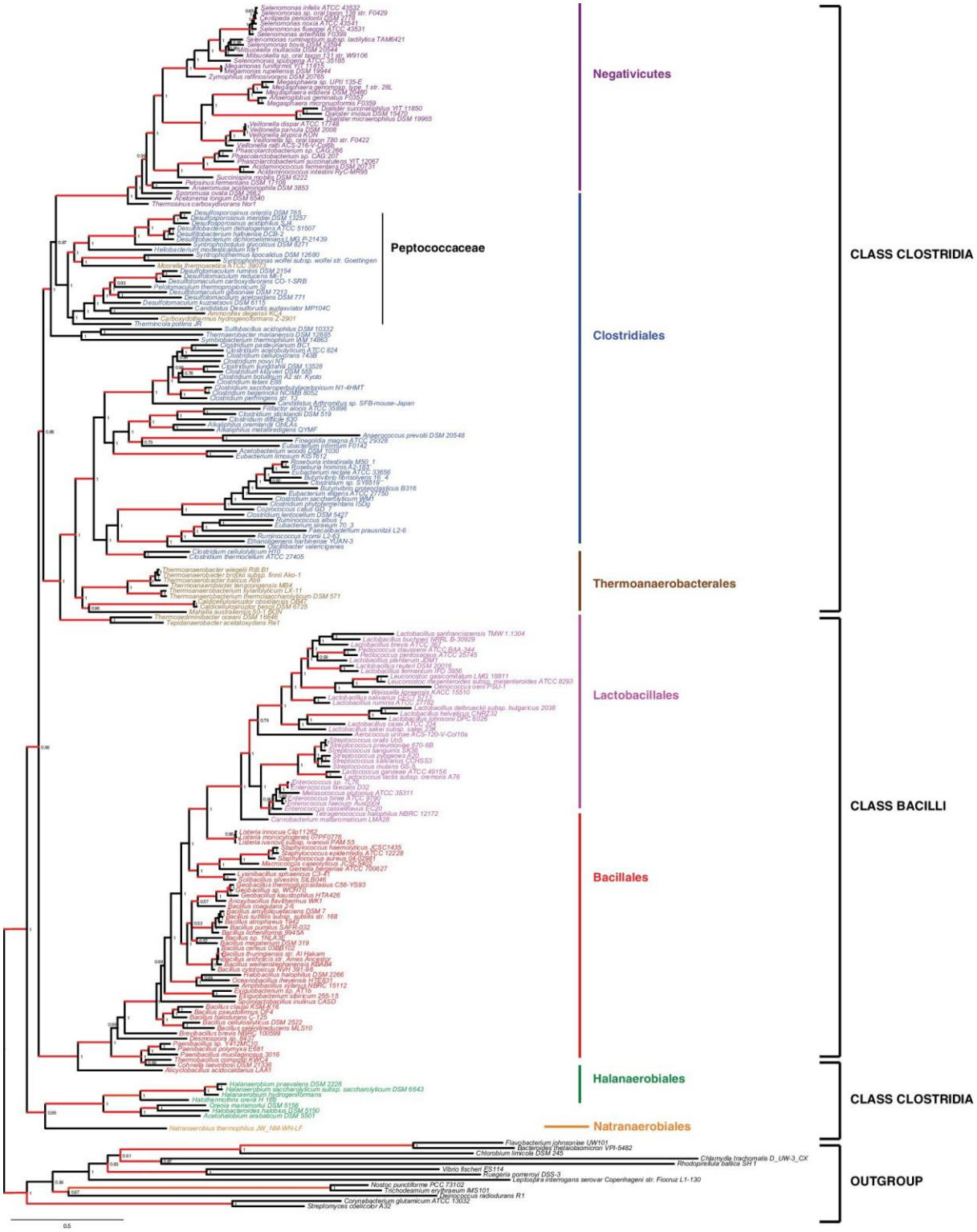

Figure A.1. Results of IC congruence test for the 47 ribosomal proteins.
IC values were mapped onto the ribosomal protein concatenation phylogeny shown in Figure 2 of (Antunes et al., 2016). Branches in red indicate congruence among markers according to IC tests. Raw results of the test are provided as Additional Data in Dryad.

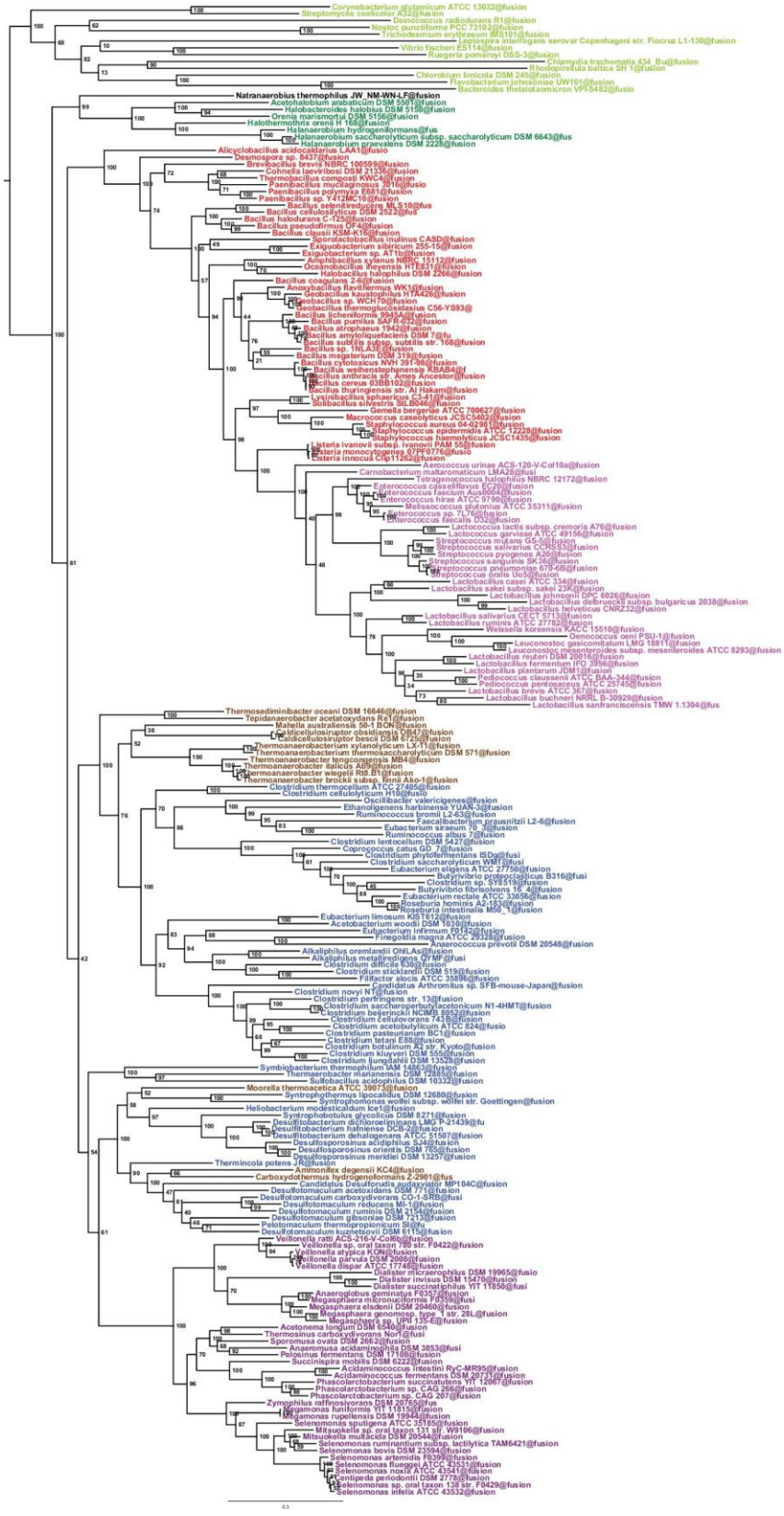

Figure A.2. Maximum likelihood phylogeny of the Firmicutes.
The tree was obtained by PhyML 3.0 based from the same concatenation of 47 orthologous ribosomal proteins as the Bayesian tree in Figure 2 and the LG+Γ4 model. Values at nodes represent non-parametric bootstrap values calculated on 100 replicates of the original dataset. The scale bar represents the average number of substitutions per site.

Hypothesis	P value
H_0 N_0	0.775
H_1	0.042
H_2	1.00E-04
H_3	1.00E-05
H_4	9.00E-05
H_5	4.00E-07
H_6	9.00E-06
N_1	0.465
N_2	0.081
N_3	0.023
N_4	0.001
N_5	1.00E-06
N_6	1.00E-04

Results from approximate unbiased test on 12 topologies alternative to the Bayesian tree in Figure 2 (hypothesis H_0N_0). P values < 0.05 are in red and indicate topologies that are significantly different from the original hypothesis H_0 N_0. Alternative topologies testing the movement of the Halanaerobiales clade are indicated by $H_{1\text{-}6}$ while those referring to movements of Negativicutes are indicated by $N_{1\text{-}6}$. These topologies are provided as Supplementary data. See text for details.

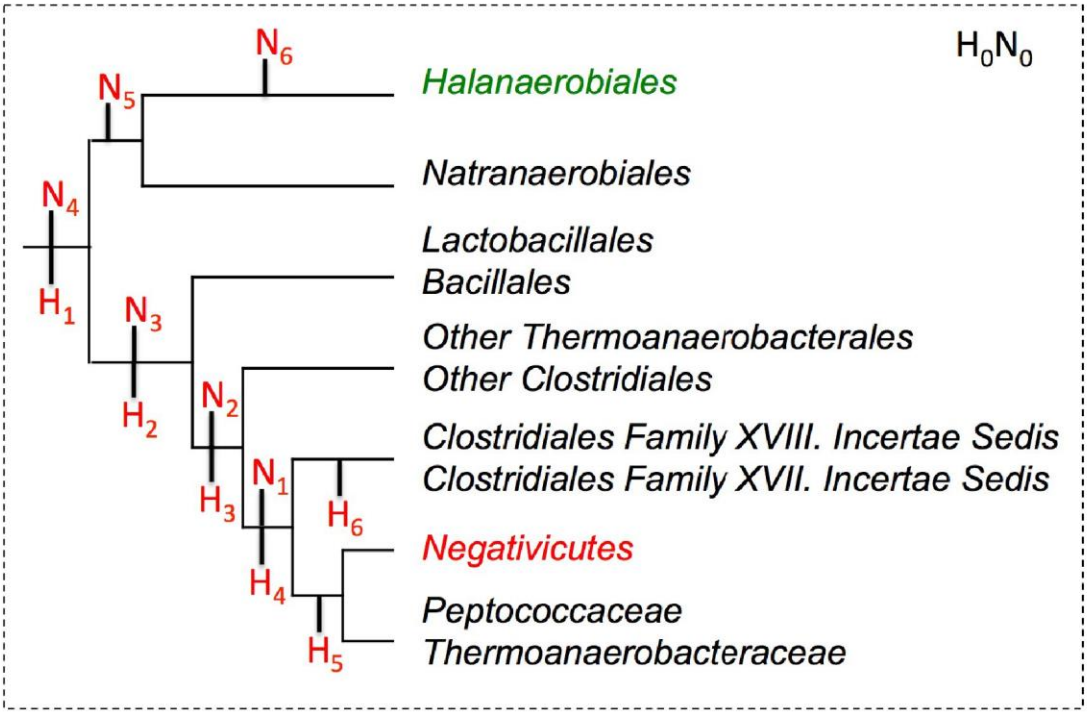

Figure A.3. Aproximate unbiased test for 12 alternate topologies.

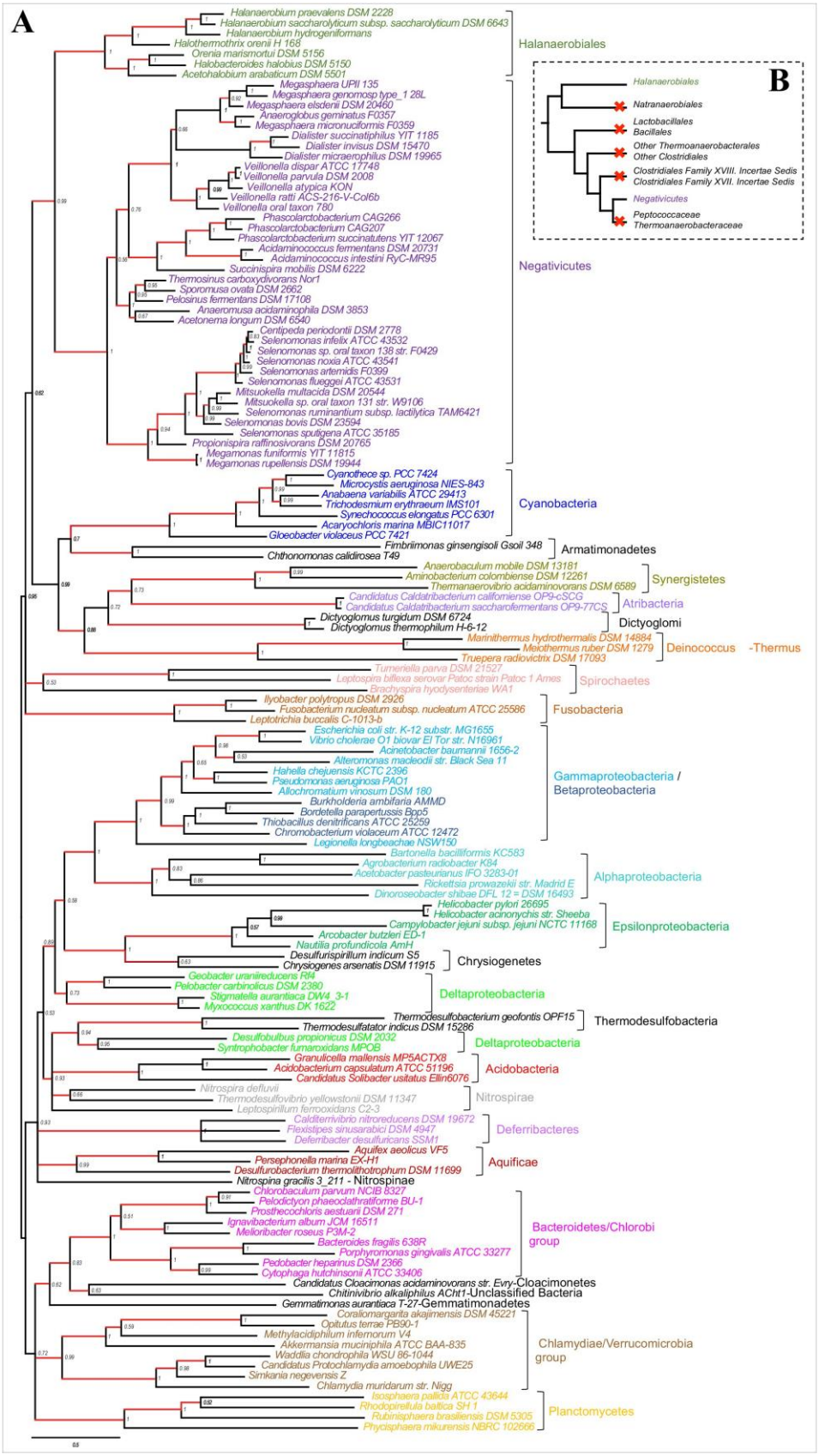

Figure A.4. Results of IC congruence test for the 4 LPS core proteins.
IC values were mapped onto the LPS core proteins concatenation phylogeny shown in Figure 3. Branches in red indicate congruence among markers (IC values>1).

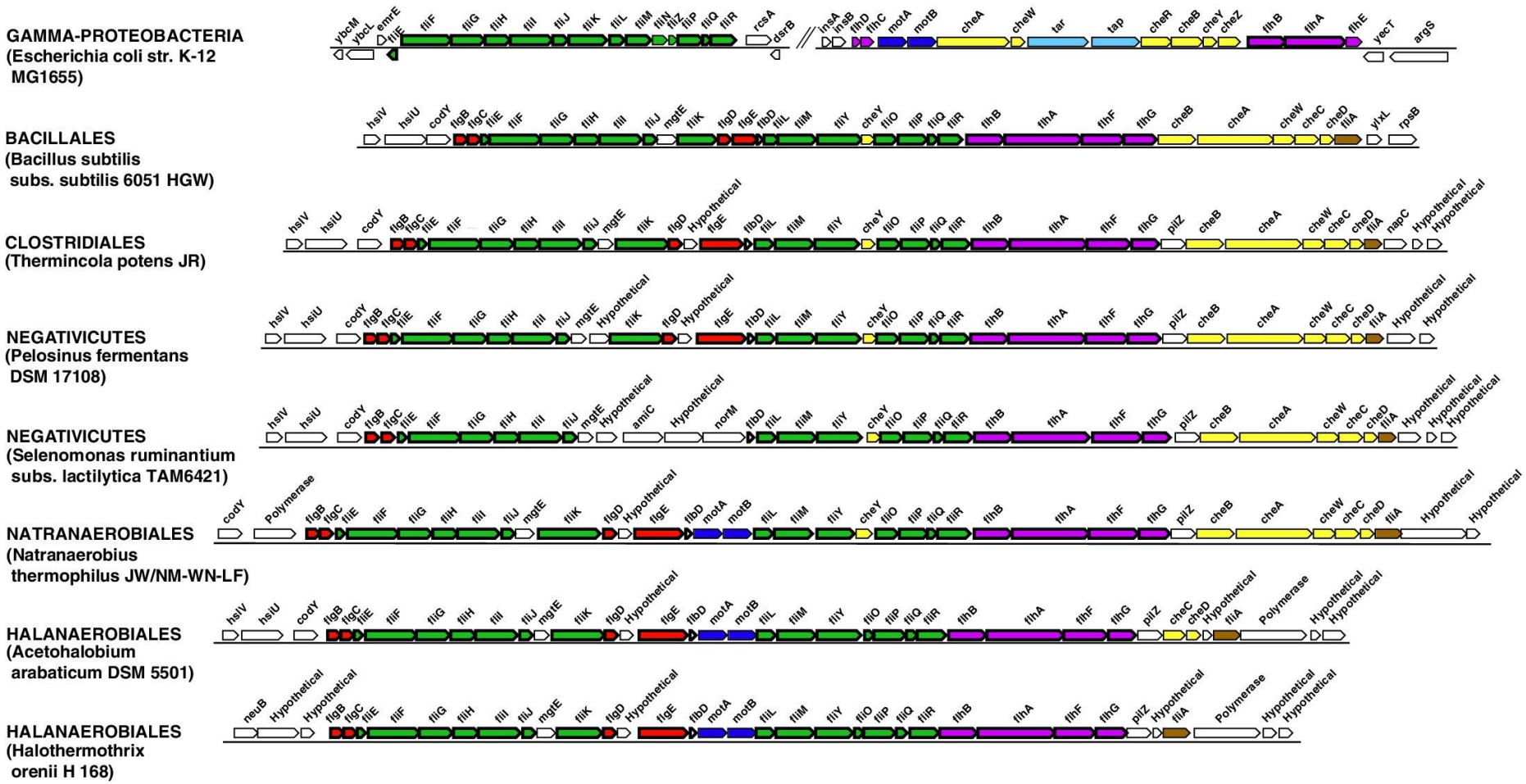

Figure A.5. Flagellar gene cluster of Negativicutes and Halanaerobiales.

Structure of the region coding for flagellar components in representative members of Negativicutes and Halanaerobiales, and its conservation with respect to their closely related monoderm relatives Therminicola potens, and Natranaerobius thermophiles, respectively. By comparison is shown the structure of the operon in Escherichia coli as representative of a classical diderm. Colors are only meant to highlight synteny.

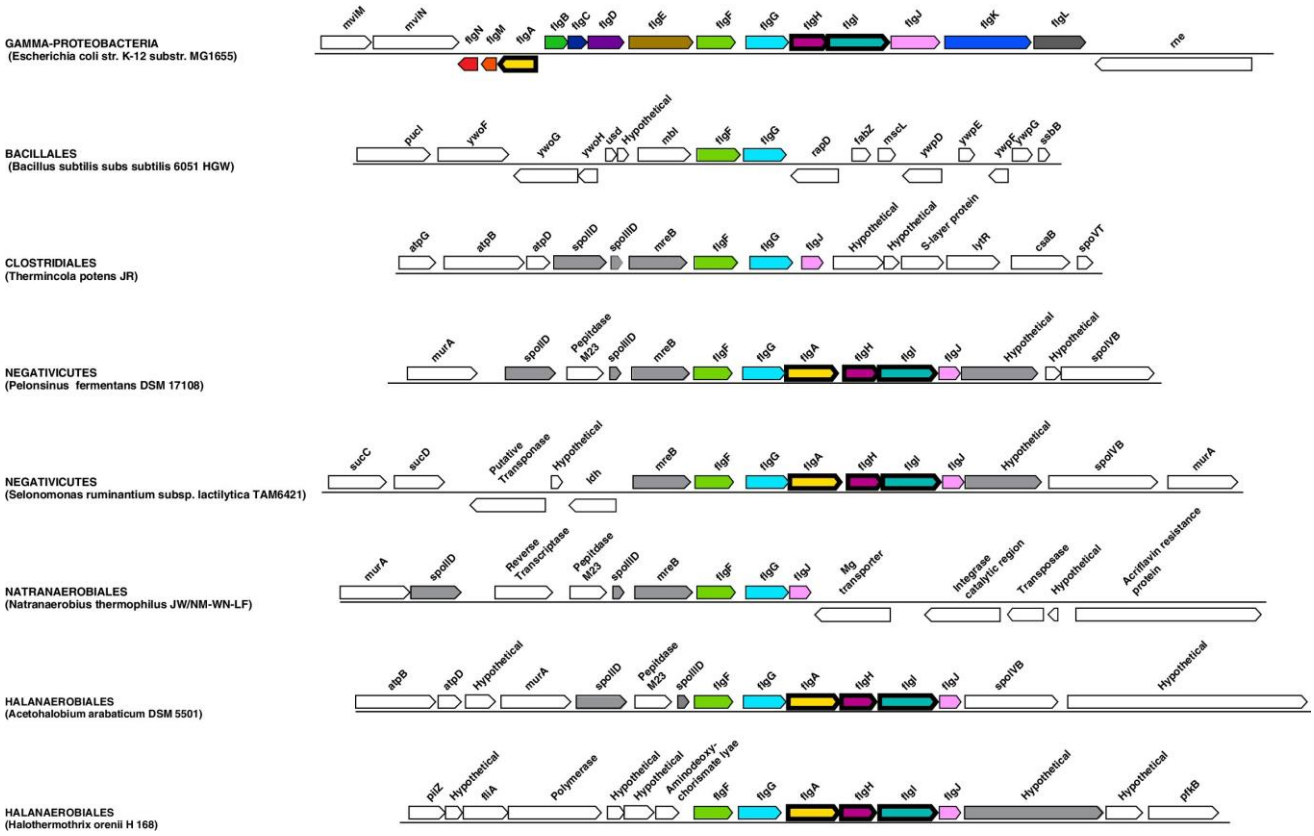

Figure A.6. Genomic context of the genes coding for flagellar rings in Halanaerobiales and Negativicutes.

Structure of the region coding for components of the flagellar rings (flgA, flgH, flgI) in representative members of Negativicutes and Halanaerobiales, in comparison with their closely related monoderm relatives Therminicola potens, and Natranaerobius thermophiles, respectively. The genomic structures in Bacillus subtilis and Escherichia coli are also shown as the most studied models for monoderm and diderm flagella. Colors are only meant to highlight synteny.

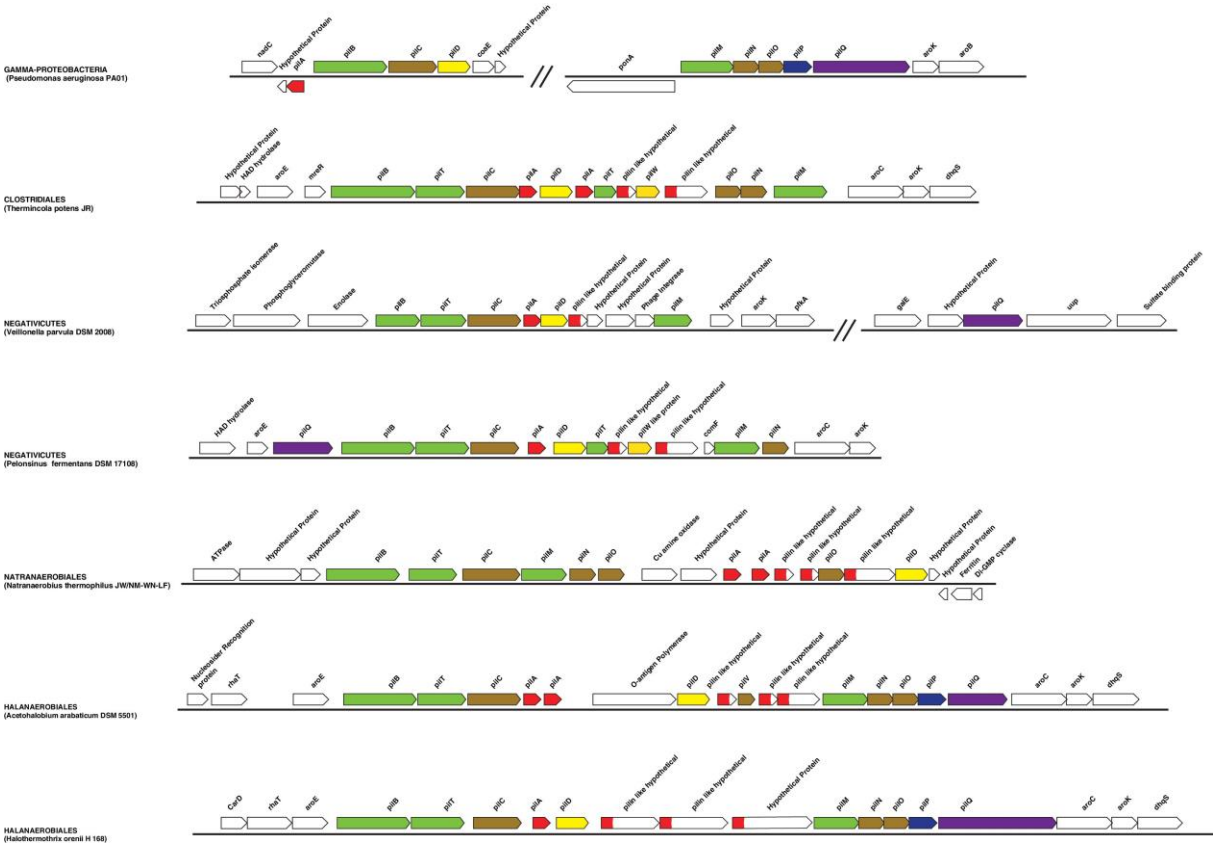

Figure A.7. Structure of the main Type IV pilus cluster in Negativicutes and Halanaerobiales.
Structure of the region coding for components of the type IV pilus in representative members of Negativicutes and Halanaerobiales, in comparison with their closely related monoderm relatives Therminicola potens, and Natranaerobius thermophiles, respectively. The genomic structure in Pseudomonas aeruginosa is also shown as the most studied model for diderm type IV pili. Colors are only meant to highlight synteny

1.2 Supplemental Figures from Poppleton et al. 2017

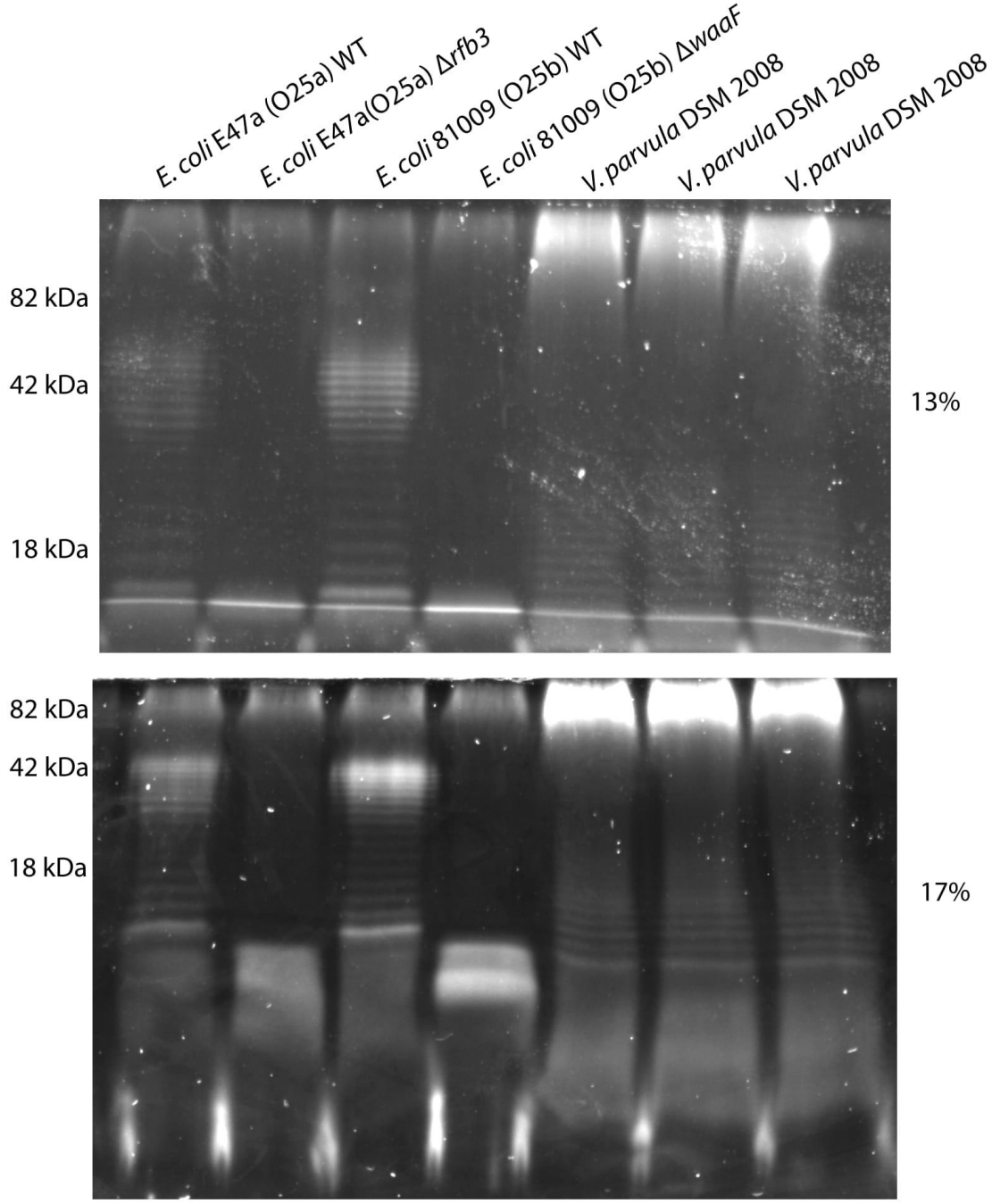

Figure A.8. Glycostain of LPS.
Thirteen and seventeen percentage SDS-PAGE loaded with identical samples and visualized with Pro-Q emerald glycostain. O-antigen producing positive control E. coli were loaded in lanes 1 and 3, while O-antigen negative extractions were loaded in lanes 2–4. Three biological replicates of V. parvula were loaded in lanes 5–7.

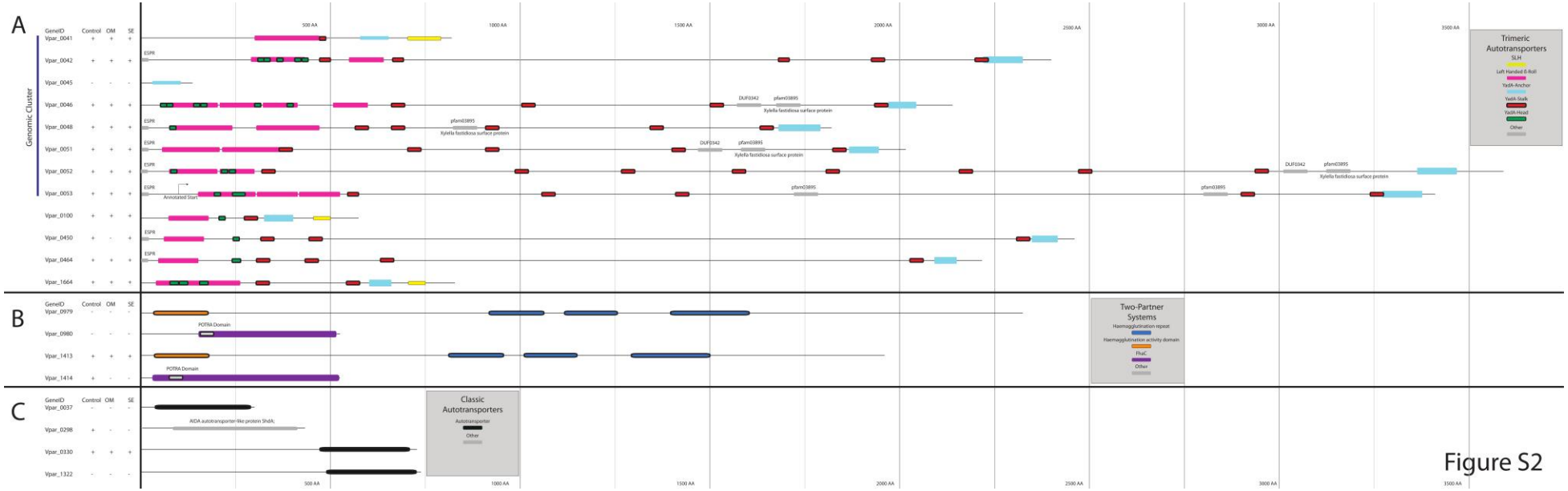

Figure A.9. Adhesins domain structure.
Figure demonstrating the domain structure of all trimeric autotransporters (A), two partner systems (B), and autotransporters (C).

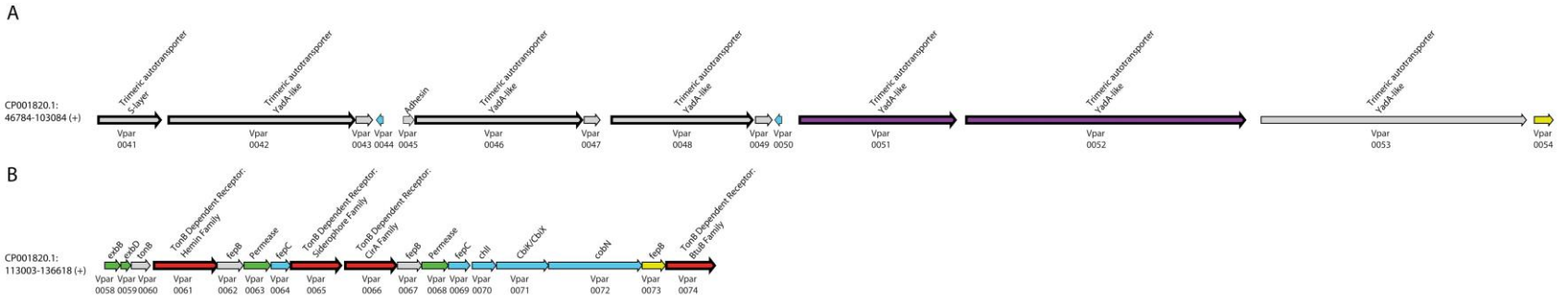

Figure S3

Figure A.10. Trimeric autotransporter (A) and TonB (B) genomic clusters.
Bold arrows represent peptides detected in the OM. Localization is presented by color: Gray, Unclear; Blue, Cytoplasmic; Green, IM; Yellow, Periplasmic; Purple, Secreted; and Red, OM.

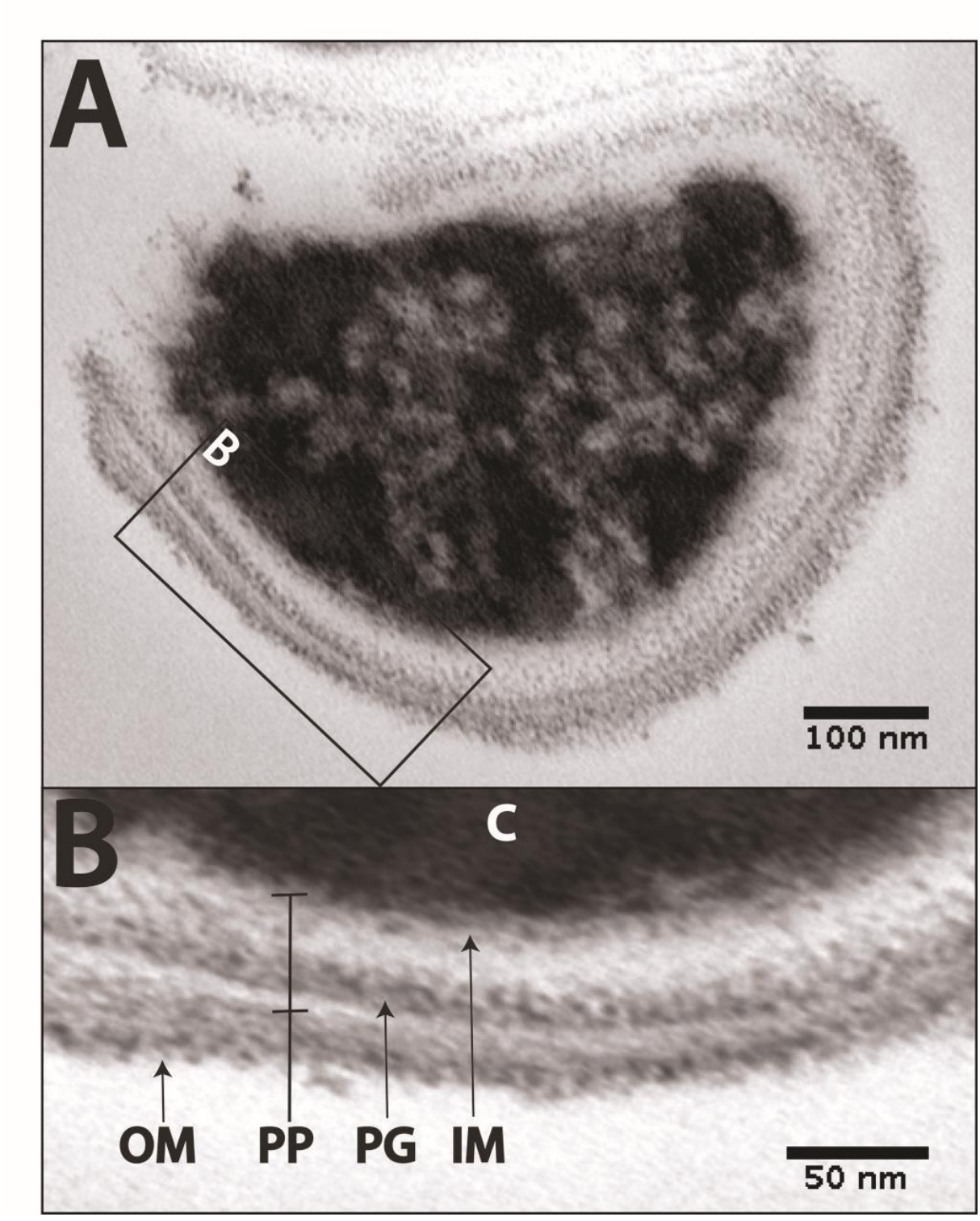

Figure A.11. High-pressure frozen cell illustrating LPS.
The ultrathin section of a high-pressure frozen cell shows the cell wall of V. parvula (A). Beside the inner membrane (IM), the outer membrane (OM), and the peptidoglycan (PG) within the periplasm, the fluffy outer leaflet of the outer membrane is visible at higher magnification (B). This might represent the LPS moiety. Bold arrows represent peptides detected in the OM

TABLE OF FIGURES

A Tale of Two Membranes: Envelope transition in Firmicutes

Abstract: :

The bacterial cell envelope is of the utmost importance in bacterial physiology, constituting the selective barrier of the bacterial cell. Curiously, two distinct types of cell envelope exist in bacteria: monoderm (Gram-positive) with a single membrane, and diderm (Gram-negative) with two membranes. When and how the transition between monoderm and diderm bacteria occurred is one of the greatest questions in evolutionary biology, and incites intense debate. Fortunately, we have ideal model organisms that can help better understand this transition: Negativicutes and Halanaerobiales. These two distinct and diverse clades of bacteria represent an evolutionary enigma; they belong phylogenetically to the classical monoderm Firmicutes, yet possess outer membranes (OM) with lipopolysaccharides (LPS) similar to classic diderm bacteria. The three goals of this Doctoral work are to: **1. Describe the outer membranes of diderm Firmicutes, 2. Elucidate the evolutionary history of these envelopes, 3. Find a mechanism of transition between diderm and monoderm Firmicutes.**

1. To characterize the cell envelopes of diderm Firmicutes we conducted a thorough bioinformatics analysis of all 1844 predicted proteins from the genome of *Veillonella parvula* DSM 2008 using eleven different localization prediction programs. These results were complemented by protein extraction with surface exposed protein tags and subcellular fractionation, both of which were analysed by liquid chromatography-tandem mass spectrometry. The merging of proteomics and bioinformatics results allowed identification of 78 OM proteins. These experimentally prove the existence of a number of outer membrane systems including an atypical Bam/Tam system for OM protein biogenesis, a complete LPS transport complex, and a large number of hypothetical proteins.

We also identified OmpM in the outer membrane of *V. parvula*. This primary mechanism of outer membrane attachment in Negativicutes is drastically different from that of *E. coli*, which uses Braun's lipoprotein (Lpp). We, therefore, investigate the distribution of these two systems in bacteria along with Tol-Pal, a less studied tether. Surprisingly, Lpp is restricted to a few γ-Proteobacteria, while OmpM is distributed among most basal bacteria, and Tol-Pal is present in the remainder. Together, this shows the presence of two different and mutually exclusive mechanisms of OM attachment in Bacteria, OmpM and Tol-Pal, with Lpp being a recent invention.

2. To elucidate the evolutionary history of Firmicute cell envelope, we undertook a thorough phylogenomic analysis. In a reconstructed phylogeny of the Firmicutes, the Negativicutes and the Halanaerobiales formed distinct branches with closely related monoderm lineages. A phylogenetic tree of the conserved LPS genes demonstrates vertical inheritance. These results strongly suggest that the OM of Negativicutes and Halanaerobiales did not originate from horizontal gene transfer, rather it was inherited from a diderm ancestor. This means that monoderm Firmicute lineages arose through multiple independent losses of the OM.

3. We hypothesized that such losses were due to the loss of a peptidoglycan-remodelling enzyme that resulted in unregulated outgrowth of peptidoglycan, and a rupture of the OM. To test this hypothesis, a survey of peptidoglycan hydrolysing enzymes was completed within the Firmicutes. A hydrolase was identified that was absent in monoderm Firmicutes, yet present in all diderm Firmicutes. We attempted insertional mutagenesis and in frame deletion of the hydrolase encoding gene in *V. parvula*. Neither was successful, suggesting that the gene was essential in our growth conditions.

Taken together, these results provide precious information on the biogenesis, functioning, and origin of the diderm cell envelopes of Negativicutes, Halanaerobiales, monoderm Firmicutes and all Bacteria. They open the way to further characterization of the outer membrane of these interesting Firmicutes and to a better understanding of the diderm to monoderm transition.

A Tale of Two Membranes: Envelope transition in Firmicutes

Résumé :

L'enveloppe des bactéries est de la plus haute importance pour leur physiologie, constituant la barrière sélective de la cellule bactérienne. Curieusement, il existe deux types distincts d'enveloppes cellulaires chez les bactéries : les monodermes (Gram-positives), avec une seule membrane, et les didermes (Gram-négatives), qui en possèdent deux. Quand et comment s'est produite la transition entre bactéries monodermes et didermes sont parmi les plus grandes questions en Biologie de l'évolution, et sources d'intenses débats. Nous disposons toutefois d'organismes modèles idéaux pour mieux comprendre cette transition : les Négativicutes et les Halanaérobiales. Ces deux clades bactériens, distincts et variés, représentent une énigme évolutionnaire. Phylogénétiquement, ils appartiennent tous deux à la famille des Firmicutes monodermes typiques. Pourtant, ils possèdent également une membrane externe avec des lipopolysaccharides (LPS), de façon similaire aux bactéries didermes. Mon travail doctoral a ainsi été orienté sur trois objectifs : **1 - Décrire les membranes externes des Firmicutes didermes ; 2 - Elucider l'histoire évolutionnaire de ces enveloppes ; 3 - Trouver un mécanisme de transition entre Firmicutes monodermes et didermes.**

1. Afin de caractériser l'enveloppe cellulaire des Firmicutes didermes, nous avons procédé à une analyse bioinformatique approfondie de l'ensemble des 1844 protéines prédites dans le génome de Veillonella parvula DSM 2008, au moyen de onze logiciels de prédiction de localisation différents. Les résultats obtenus ont été complétés par extraction protéique avec des marquages de protéines exposées en surface, ainsi que par fractionnement sous-cellulaire ; dans les deux cas, l'analyse a été effectué via chromatographie liquide couplée à la spectrométrie de masse en tandem. Le recoupement des résultats de bioinformatique avec ceux de protéomique a permis d'identifier 78 protéines associées à la membrane externe. Il a ainsi été possible de prouver expérimentalement la présence de certains systèmes relatifs à la membrane externe, notamment un système Bam/Tam atypique pour la biogénèse de protéines de membranes externes, un complexe de transport de LPS complet, ainsi qu'un grand nombre de protéines hypothétiques.

Nous avons également pu identifier OmpM dans la membrane externe de V. parvula. Ce principal mécanisme d'attachement de la membrane externe chez les Négativicutes est drastiquement différent à celui d'E. coli, qui utilise des lipoprotéines de Braun (Lpp). Nous avons donc analysé la distribution de ces deux systèmes chez les Bactéries, ainsi que du système Tol-Pal, une ancre moins étudiée. De façon surprenante, les Lpp sont limités à quelques γ-Protéobactéries, alors qu'OmpM est présent chez la plupart des bactéries basales et Tol-Pal chez les bactéries restantes. Ceci démontre l'existence de deux mécanismes différents et mutuellement exclusifs, OmpM et Tol-Pal, pour l'attachement de la membrane externe chez les bactéries, faisant de Lpp une invention récente.

2. Pour élucider l'histoire évolutionnaire de l'enveloppe cellulaire des Firmicutes, nous avons entrepris une analyse phylogénomique approfondie. Dans une phylogénie reconstruite des Firmicutes, les Négativicutes et les Halanaérobiales forment des branches distinctes avec les proches lignées monodermes. Un arbre phylogénétique des gènes conservés codants pour le LPS démontre une transmission verticale. Ces résultats suggèrent fortement que la membrane externe des Négativicutes et des Halanaérobiales n'a pas pour origine un transfert latéral de gènes, mais a plutôt été hérité depuis un ancêtre diderme. Cela signifierait ainsi que les différentes lignées monodermes des Firmicutes seraient apparues suite à plusieurs pertes indépendantes de la membrane externe.

3. Nous avons émis l'hypothèse que de telles pertes étaient dues à la perte d'une enzyme de modification du peptidoglycane, résultant en une surcroissance non-régulée de ce dernier jusqu'à la rupture de la membrane externe. Afin de tester notre hypothèse, nous avons étudié la présence au sein des Firmicutes d'enzymes hydrolysant le peptidoglycane. Nous avons pu détecter une hydrolase, absente chez les Firmicutes monodermes, mais présente chez tous les Firmicutes didermes. Nous avons alors tenté des mutagénèses insertionnelles et des délétions du gène encodant l'hydrolase chez V. parvula. Aucune des deux approches ne réussît, ce qui suggère que ce gène est essentiel dans nos conditions de culture.

Dans leur ensemble, mes résultats fournissent de précieuses informations sur la biogénèse, le fonctionnement, et l'origine des enveloppes cellulaires des didermes Négativicutes et Halanaérobiales, des Firmicutes monodermes, ainsi que de l'ensemble des bactéries. Ils ouvrent ainsi la voie à de futures caractérisations de la membrane externe de ces Firmicutes intéressants, et à une meilleure compréhension de la transition diderme/monoderme.