Genome-wide analysis of the Firmicutes illuminates the diderm/monoderm transition

Najwa Taib, Daniela Megrian, Jerzy Witwinowski, Panagiotis Adam, Daniel Poppleton, Guillaume Borrel, Christophe Beloin, Simonetta Gribaldo

Genome-wide analysis of the Firmicutes illuminates the diderm/monoderm transition

Najwa Taib[1,2#], Daniela Megrian[1,3#], Jerzy Witwinowski[1], Panagiotis Adam[1^], Daniel Poppleton[1&], Guillaume Borrel[1], Christophe Beloin[4], and Simonetta Gribaldo[1]

[1] Unit Evolutionary Biology of the Microbial Cell, Department of Microbiology, Institut Pasteur, UMR 2001 CNRS, Paris, France

[2] Hub Bioinformatics and Biostatistics, Department of Computational Biology, Institut Pasteur, USR 3756 CNRS, Paris, France

[3] Sorbonne University, Collège doctoral, F-75005 Paris, France

[4] Unit Genetics of Biofilms, Department of Microbiology, Institut Pasteur, Paris, France

[^] Current affiliation: Environmental Microbiology and Biotechnology, Faculty of Chemistry, University Duisburg-Essen, Germany

[&] Current affiliation: Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, UK

# these authors contributed equally to this work

* Correspondence: simonetta.gribaldo@pasteur.fr

Keywords: Limnochordia; Cell Envelope; Outer Membrane; LPS; Phylogenomics; Evolution; Diderm Firmicutes

Summary

The transition between cell envelopes with one membrane (Gram-positive or monoderm) and those with two membranes (Gram-negative or diderm) is a fundamental open question in the evolution of Bacteria. The evidence of two independent diderm lineages, the Halanaerobiales and the Negativicutes, within the classically monoderm Firmicutes has blurred the monoderm/diderm divide and specifically anticipated that other members with an outer membrane (OM) might exist in this phylum. Here, by screening 1,639 genomes of uncultured Firmicutes for signatures of an OM, we highlight a third and deep branching diderm clade, the Limnochordia, strengthening the hypothesis of a diderm ancestor and multiple transitions leading to the monoderm phenotype. Phyletic patterns of over 176,000 protein families constituting the Firmicutes pan-proteome identify those that are specific to the three diderm lineages, and suggest new potential players in OM biogenesis. In contrast, we find practically no largely conserved core for monoderms, a fact possibly linked to different ways of adapting to OM loss. Phylogenetic analysis of a concatenation of main OM components totalling nearly 2000 amino acid

positions illustrates the common origin and vertical evolution of most diderm bacterial envelopes. Finally, mapping the presence/absence of OM markers onto the tree of Bacteria highlights the overwhelming presence of diderms and the non-monophyly of monoderms, pointing to an early origin of two-membraned cells and the derived nature of the Gram-positive envelope following independent OM losses.

Introduction

The cell envelope is one of the most essential and ancient features of life; yet, most aspects of its evolutionary history remain obscure. In particular, it is unclear how cell envelopes with two membranes (Gram-negative or diderms) and those with one membrane (Gram-positive or monoderms) came into being, and how such important transition occurred[1–5].

The Firmicutes are the textbook example of Gram-positive bacteria, but surprisingly include two clades, the Negativicutes and the Halanaerobiales, that display an outer membrane (OM) with lipopolysaccharide (LPS)[6–9]. We recently showed that they form two phylogenetically distinct lineages, each close to different monoderm relatives within the Firmicutes. In addition, phylogenetic analysis of core biosynthetic LPS genes indicated that these were not acquired through horizontal gene transfer (HGT) from other diderm bacteria. Finally, identification and annotation of putative OM markers in the genomes of Halanaerobiales and Negativicutes suggested that these two lineages display unique cell envelopes with specific characteristics[10], such as for example the mechanism of OM attachment, which is different from that of *Escherichia coli*[5,11]. These bioinformatics predictions were confirmed by characterizing the first OM proteome from the model diderm Firmicute *Veillonella parvula*[12]. From these results, we put forward the hypothesis that a diderm envelope with LPS was already present in the ancestor of all Firmicutes and was retained in the Negativicutes and Halanaerobiales while it was lost multiple times independently during the diversification of this phylum to give rise to the classical Gram-positive cell envelope architecture[10]. This hypothesis specifically anticipated that other diderm lineages may exist in the Firmicutes.

Recently, 1,639 new genomes from uncultured Firmicutes obtained from metagenomes were made available[13], providing an exceptional wealth of new data to explore. Here, we analyzed these genomes to investigate the presence of additional diderm members and to obtain further information on the diderm/monoderm transition in this phylum. Our results strengthen the hypothesis of a diderm ancestor of the Firmicutes and possibly all Bacteria, and the derived nature of the monoderm cell envelope.

Results

Screening uncultured Firmicutes genomes for OM markers highlights a third diderm lineage

We retrieved 1,639 genomes annotated as Firmicutes from the Uncultured Bacteria and Archaea (UBA) dataset of Parks et al.,[13]. These genomes were isolated from different environments and their quality goes from partial to near complete (Supplementary Table 1 and Supporting Data). 514 UBA genomes having less than 35 ribosomal proteins were considered as too partial and discarded from further analysis. To robustly place these new genomes in the reference Firmicutes phylogeny, we included the 1,125 UBA genomes in a phylum-level tree together with 230 representatives of all Firmicutes families, and members of major bacterial phyla as outgroup (Methods).

The resulting maximum-likelihood (ML) tree is well resolved (Figure 1 and Supporting Data). The UBA genomes significantly enrich the genomic coverage of the Firmicutes, by spanning the entire diversity of this phylum, in particular the Clostridia. Most UBA genomes fall into known Firmicutes families consistently with their inferred taxonomy[13], whereas other fall into clades that do not contain any genome representative of known families, and were given a loose taxonomic assignment[13] (Figure 1). Only one UBA genome belongs to the Halanaerobiales, possibly a bias due to the difficulty of assembling these GC-rich genomes or a poorer sampling from the environments where they thrive. In contrast, these new genomes significantly enrich the coverage for Negativicutes (39 UBAs). The deep branching of Halanaerobiales and the placement of Negativicutes within Clostridia are both well supported and are consistent with previous analyses[6,10,14]. In particular, Antunes et al. specifically tested alternative branchings of the Halanaerobiales by approximately unbiased (AU) tests and showed that they were all strongly rejected by the data[10]. Interestingly, while the monoderm lineage *Natranaerobius thermophilus* branched with Halanaerobiales in previous analyses[10,14], it now groups with *Dethiobacter alkaliphilus* and 17 other UBA genomes at the base of Bacilli (Figure 1). The increased genomic coverage likely enhanced the phylogenetic signal and helped to correctly place these lineages, which is more consistent with their monoderm phenotype.

To investigate the existence of additional diderm lineages among the UBA Firmicutes, we screened them for the presence of four conserved genes for LPS biosynthesis and six protein domains previously used as markers of Gram-negativity[7] (Methods). Consistently with our previous study[10], we found homologues of these OM markers in all Negativicutes and Halanaerobiales UBA genomes, but surprisingly we also found them in 46 unclassified UBA genomes belonging to a particularly interesting clade (Supplementary Table 2). It represents the second deepest-branching group in the reference phylogeny after the Halanaerobiales (Figure 1), and includes a single isolated member, *Limnochorda pilosa*, identified from a brackish meromictic lake, and defining the recently proposed class Limnochordia[15]. The recently sequenced *L. pilosa* genome revealed the presence of classical OM markers consistent with microscopy evidence for a diderm cell envelope[16]. Most UBA genomes that we assign to Limnochordia were assembled from anaerobic mud and digester samples (Supplementary Table 1) and form a diverse group (Supplementary Figure 1). These results show that Limnochordia represent a third and deep-branching diderm lineage within the Firmicutes, strengthening the hypothesis of a diderm ancestor.

Distribution of protein families and domains reveals the functional core of diderm Firmicutes

The existence of three distinct diderm lineages embedded in the classical monoderm Firmicutes provides an exceptional opportunity to gain insights into the diderm/monoderm transition by analyzing the distribution of proteins across this phylum. To this aim, we assembled a reduced genome databank of 316 representative Firmicutes (including the newly identified UBA diderm Firmicutes): 47 Limnochordia, 101 Negativicutes and 17 Halanaerobiales, for a total of 165 diderm and 151 monoderm taxa (Methods). We used this databank to compare the proteomes of monoderm and diderm lineages across the phylum. We calculated the pan-proteome of Firmicutes as composed of 176,024 protein families, 15,964 being present in at least five genomes and which were thus kept for further analysis (Methods).

We then built a presence/absence matrix and calculated the distribution of these 15,964 protein families (Figure 2). Some families are specific to Halanaerobiales (327), to Negativicutes (1,820), or to Limnochordia (2,080) (Figure 2a and Supplementary Table 3). However, these might include innovations specific to these clades or linked to their phylogenetic placement and not necessarily to the diderm phenotype. Of the 15,964 families, 131 are shared uniquely between Halanaerobiales and Limnochordia, 73 between Limnochordia and Negativicutes, and 26 between Negativicutes and Halanaerobiales. Finally, 41 protein families are present in at least one member of each of the three diderm Firmicute lineages but are totally absent from monoderm Firmicutes, and will be referred to hereafter as the diderm "strict core" (Figure 2a, in bold). Consistently, these families include components of known OM systems (Supplementary Table 3) such as LpxABD (LPS synthesis), BamA (OM biogenesis) and FlgHI (Flagellar rings). However, some false negatives may arise from this approach. This is for example the case of LpxC and KdsABD (keto-deoxyoctulosonate (KDO) synthesis), which do not appear in the diderm strict core because homologues are present in a few monoderm genomes (Figure 3 and Supporting Data). In contrast, this analysis did not identify a strict core of monoderms, because, within the 3,863 protein families totally absent from diderm Firmicutes, only three are present in > 50% of monoderm genomes (Supplementary Table 3, highlighted in grey). Therefore, to relax our criteria, we implemented a complementary approach based on hierarchical clustering (HC) on the 15,964 protein families (Methods). Among the 3,500 clusters generated by the HC approach, four (HC5, HC7, HC8 and HC9) appeared particularly enriched in the three diderm clades with respect to their monoderm relatives and contained at least one of the previously identified "strict core" diderm families present in more than 40% of diderm Firmicutes. These four clusters total 120 protein families (Supplementary Table 4), 95 of which were also retrieved with an alternative clustering approach (Methods, Supplementary Table 5). The HC5 (14 protein families), HC7 (67 protein families), and HC9 (8 protein families) clusters show variable distribution patterns across diderm Firmicutes and appear to be mostly restricted to the Negativicutes (Figure 2b and Supplementary Table 4). In contrast, cluster HC8 has the sharpest pattern of enrichment in all three diderm Firmicutes clades (Figure 2b). It includes 31 protein families, which comprise some of the OM markers that were missed by the strict presence/absence criterion, such as LpxC and KdsABD (Supplementary Table 4), confirming the highest sensitivity of the clustering approach. Consistently with the strict core analysis, here again we could not highlight any cluster specifically enriched in monoderms with respect to diderms (Supplementary Data).

Finally, we used a third approach based on Pearson Correlation Coefficient (PCC) (Methods). We found that 83 families correlate with the diderm phenotype (PCC >= 0.5) (Figure 2c). Of these, 26 are present in HC8, 56 are totally absent in monoderm Firmicutes, and 35 are present in all three diderm Firmicutes lineages (Supplementary Table 6). Some important OM markers were nevertheless not recovered by the analysis of protein families. This is for example the case of TamB, an inner membrane (IM) protein involved in the insertion of proteins in the OM. Another example is that of OmpM, a porin responsible for tethering the OM in Negativicutes, and some of the components of the Lpt system to export LPS to the OM (LptB and LptD). These homologues are not sufficiently conserved at the sequence level and ended up being split among different families. We therefore complemented the protein family approach by one based on protein domains (Methods). We identified all Pfam domains in our Firmicutes databank, and grouped proteins using three different approaches (Methods). We then computed the PCC of each of these groups with the diderm phenotype (Supplementary Figure 2 and Supplementary Table 7). This allowed to identify OM markers that were missed by the protein family approach, such as TamB and OmpM. All results (protein families and Pfam domains), were merged. For the proteins identified by more than one method, we kept the corresponding family or domain presenting the best PCC. Finally, we only kept the protein families or domains present in at least one member of the three diderm Firmicutes lineages. These analyses led to identify 52 protein families/domains that are specifically enriched in the three diderm Firmicutes lineages, the majority of which are either totally or mostly absent from monoderm Firmicutes and include crucial functions linked to the cell envelope and the OM (Figure 3a). Interestingly, they also include putative functions not immediately linked to OM biogenesis, such as RuvC (endonuclease) and RecX (a regulator of RecA, involved in DNA repair). Other proteins have only a generic annotation (e.g. Peptidase_S55, Peptidase_M23, AMIN) and might be linked to peptidoglycan (PG) related functions. Finally, some proteins are totally uncharacterized. For example, DUF3084 which is among the most highly correlated with the diderm phenotype, and is practically absent in monoderms. Finally, we identified 51 protein families and Pfam domains that correlate with the monoderm phenotype (Figure 3b, Supplementary Figure 2, Supplementary Table 8). However, and consistently with the strict core analysis, most of these proteins/domains are not widely distributed across monoderms, even those commonly assumed to be markers of Gram-positive (e.g. sortase, LPXTG anchor), and only one protein family (containing the S4 domain, loosely annotated as involved in RNA binding and translation regulation) shows a PCC higher than 0.9. Therefore, unlike diderm Firmicutes, it appears that there is *de facto* no conserved core for monoderm Firmicutes.

A large OM gene cluster is a distinguishing feature of all diderm Firmicutes and reveals potential novel functions related to OM biogenesis

Very little experimental data are available on the nature of the cell envelope of diderm Firmicutes. In order to gather further information on the putative functions of the 52 family/domains most correlated with the diderm phenotype, we applied a "guilty by association" strategy by looking at their genomic environment. Twenty nine of them revealed to be part of a large gene cluster (up to 65 kb) that we previously identified in Negativicutes and Halanaerobiales[10]. This cluster, which will be hereafter referred to as the "OM cluster" is also present in all Limnochordia genomes with a very similar gene order (Figure 4) and therefore appears to be a distinguishing characteristic of all diderm Firmicutes.

The OM cluster codes for a number of important systems known to be involved in the biogenesis of the bacterial diderm cell envelope[17,18] (Figure 4). It contains the main players in the synthesis of LPS and its transfer to the OM (Figure 4, in blue). This includes the pathway responsible for synthesis of Lipid A, the innermost component of LPS (LpxACD/LpxI/LpxB/LpxK). While the enzymes for the core oligosaccharide component of LPS are present in Halanaerobiales and Negativicutes (WaaM, KdsABCD), they are absent altogether in Limnochordia, where they might be replaced by a large number of unrelated glycosyl transferases present in the cluster (Figure 4). All diderm Firmicutes seem to have a conserved LipidA flippase (MsbA), and the Lpt system for LPS transport to the OM (LptABC/LptFG/LptD)[19]. It is intriguing to note that the gene coding for LptD is not close to those encoding the other Lpt components, but frequently lies next to two genes with no annotated function or conserved domains in both Negativicutes and Halanaerobiales (Hypothetical 1 and 2 in Figure 4). This conserved three-gene arrangement suggests functional linkage and two potentially novel components of the machinery for LPS transport to the OM in diderm Firmicutes. Similarly, a gene coding for a third hypothetical protein is always present in the OM gene cluster (Figure 4, Hypothetical 3). No functional domains are annotated, but it's strong conservation within the LPS genes strongly suggest some involvement in this pathway.

The OM cluster also includes another very important system responsible for the assembly of OM proteins (OMPs) into the lipid bilayer (Figure 4, in orange). It requires a coordinated process of folding into a β-barrel structure and membrane integration and it is accomplished by the β-barrel assembly machinery (BAM) complex whose main component is BamA (Omp85)[17]. In *Escherichia coli* and other Proteobacteria, an additional complex known as the translocation and assembly module (TAM) is present[20]. All three diderm Firmicutes lineages seem to possess a single system for OMP biogenesis composed of TamB and BamA, and one or multiple copies of the periplasmic chaperone Skp (Figure 4). This TamB/BamA complex was previously proposed to constitute an ancestral configuration predating the emergence of the Bam and Tam machineries in the evolution of Bacteria[10,20]. The TamB homologues of Limnochordia are longer than their counterparts in Halanaerobiales and Negativicutes and frequently have an N-terminal extension with no identifiable domains. This might indicate the existence of yet unknown novel functions or interactions of this machinery in diderm Firmicutes.

Curiously, the genome of *L. pilosa* and some other uncultured members of Limnochordia do not possess a homologue of the gene coding for TamB (Figure 4 and Supplementary Figure 3). It is unclear at this stage if these are true absences or repeated assembly errors.

One or multiple copies of the genes coding for OmpM can be observed in the OM cluster (Figure 4, in red). OmpM is a porin with SLH domains that has been shown to attach non covalently the OM to a modified PG in Negativicutes, constituting an atypical OM tethering system that is different from the well-known Braun's lipoprotein [11]. This suggests that diderm Firmicutes may all use a similar strategy to attach their OM to PG. Curiously, in Halanaerobiales and Limnochordia, one of the *ompM* genes lies in a conserved context including homologues of genes coding for an ExbBD/TonB machinery, responsible for the active transport of molecules to the OM by exploiting the energized IM[21]. Moreover, these genes lie together with a gene coding for cohesin, a protein generally involved in DNA repair and gene regulation (Figure 4). Such strong synteny conservation may indicate a functional link among these proteins which is intriguing and will need experimental verification.

In Negativicutes only, the OM cluster contains a conserved four-gene arrangement coding for the IM components of the Mla machinery (MlaEFD) and the OM channel TolC (Figure 4, in green). The Mla system is responsible in diderm bacteria for maintaining lipid bilayer asymmetry and OM barrier by the transport of phospholipids from and to the IM[22,24]. No clear homologues of the periplasmic and OM components of the Mla system (MlaABC) could be identified in Negativicutes (Figure 4). This may suggest a novel mechanism to maintain lipid asymmetry in the Negativicutes involving TolC. Moreover, as Halanaerobiales and Limnochordia do not have any of these Mla coding gene homologues in their genomes, it remains to be understood how they cope with the absence of such crucial system for membrane integrity, or if they use a non-homologous machinery.

In flagellated diderm Firmicutes, the OM cluster also contains six genes encoding the flagellar proteins FlgFGAHIJ, including those for the specific P- and L-ring structures that in diderm bacteria serve to anchor the flagellum to the OM[23] (Figure 4, in pink).

Finally, a number of genes are highly conserved in the OM cluster in all three diderm Firmicutes clades but have only a generic annotation (Figure 4, in yellow). Some of these are included in a conserved six-genes arrangement (sometimes interrupted by flagellar genes) which codes for: a member of the S55 peptidase family, a protein annotated as N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase (NAGPA) involved in protein glycosylation; a peptidase of the M23 family, a distant homologue of the lytic transglycosylase SpoIID, a homologue of the sporulation sigma factor SpoIIID, and a MreB homologue. Interestingly, a SpoIID homologue was recently shown to be involved in cell division in *Chlamydia trachomatis* [24] and it is tempting to speculate that some of these proteins are also involved in PG remodelling/daughter cell separation in diderm Firmicutes. Two other uncharacterized genes are always next to each other in the OM cluster (Figure 4): one codes for the uncharacterized protein DUF3084, and the other for RuvC (Holliday junction resolvase). Both proteins are among most correlated with the diderm phenotype (see below, Figure 3), strongly suggesting functional interaction and a role in OM biogenesis.

Phylogenomic analysis does not support acquisition of the OM by horizontal gene transfer

The presence of the OM cluster may suggest that it was acquired by HGT. We previously showed by phylogenetic analysis of a concatenation of the four core LPS proteins (LpxABCD), that the sequences from Halanaerobiales and Negativicutes are closely related, match their reference species phylogeny, and do not stem from within another diderm bacterial phylum. We interpreted this result as support that these genes (and by extension the whole OM cluster) were not acquired via HGT, but were rather inherited from the ancestor of all Firmicutes, which would therefore have been a diderm with LPS[5,10]. Although in our opinion it is unlikely, the possibility of an acquisition in either the ancestor of Halanaerobiales or Negativicutes followed by a further HGT between the ancestors of these two clades remained open. We think that its presence in Limnochordia weakens this scenario, as this would imply an additional ancient transfer event. Nevertheless, in order to investigate further the HGT hypothesis, we searched for OM gene clusters similar to the one present in diderm Firmicutes in our Firmicutes databank as well as in 377 genomes representatives of major bacterial phyla (Methods). We could confirm that no other bacterial phylum possesses any gene cluster similar to the one in diderm Firmicutes (Supplementary Figure 3). In most diderm bacteria, the OM genes are in fact separated in a number of small clusters as in *E. coli*. Interestingly, bigger gene clusters could be observed in some diderm phyla that are evolutionarily close to the Firmicutes (Armatimonadetes and Synergistetes), but never as large as the one present in diderm Firmicutes (Supplementary Figure 3). Taken together, these results weaken the hypothesis of an acquisition of the OM in diderm Firmicutes.

An additional argument against the HGT scenario is that the OM cluster also contains genes that are not specific of diderm Firmicutes but are also present in monoderm lineages, such as *fabZ* (fatty acid metabolism), *murA* (cell wall synthesis), *mreB* (cell shape), *spoIID*, *spoIIID* (sporulation) (Figure 4). Many of these genes lie at the beginning of the OM cluster in diderm Firmicutes but are also similarly clustered in monoderm Firmicutes (Figure 4). Under the hypothesis of a diderm ancestor of all Firmicutes[10], these genes could be remnants of an ancestral OM cluster which would have been lost in monoderms. Conversely, under the triple HGT scenario, it is difficult to explain why the OM cluster would have been inserted exactly at the same genomic position three times independently in Limnochordia, Halanaerobiales, and Negativicutes.

Finally, the HGT hypothesis is not supported by phylogenetic analysis. Among the OM markers encoded in the gene clusters previously calculated, we selected the ones most widely conserved in diderm bacteria. These were gathered into a large concatenated dataset, a strategy routinely used to increase ancient phylogenetic signal[25]. We built two alternative concatenations, one comprising 11 markers (FabZ, LpxACD/LpxB/LpxK, KdsBADC, LptB) totaling 146 taxa and 1,998 amino acid positions, and a second one also including the flagellar proteins (FlgFGAHIJ) totaling 3,705 amino acid positions but less taxa (122 taxa) (Methods). A number of diderm phyla could not be included in the concatenation because they lacked more than half of these markers, or because their genomes did not have them in cluster (Supplementary Figure 6), preventing their clear identification. These datasets are larger than the LPS concatenation we analysed previously[10]. Given the small number of positions with respect to the large evolutionary distances analysed, ML trees from both concatenations are not totally resolved, but display nevertheless a topology in overall agreement with known taxonomy (Figure 5a and Supplementary Figure 4, respectively). In particular, we observe a well-supported monophyly of major bacterial diderm phyla, and an overall topology that is consistent with the known relationships within Bacteria, notably the two large clades of Terrabacteria and Gracilicutes[26,27]. These results indicate that the OM markers were present in the ancestors of each of these major diderm phyla and have not been subjected to extensive HGT during bacterial diversification. Consistently, we observe the monophyly of the three diderm Firmicutes clades, which indicates that their OM markers are more closely related among them than to other bacteria, and do not stem from within any specific diderm bacterial phylum, which would have been expected if these markers were acquired from HGT (Figure 5a). Moreover, that clade formed by diderm Firmicutes groups with other Terrabacteria phyla, in agreement with the phylogenetic placement of the Firmicutes.

Together, these results indicate that an OM with LPS was already present in the ancestor of the Halanaerobiales, Negativicutes, and Limnochordia, which is by definition the ancestor of all Firmicutes. They strengthen the diderm-first hypothesis for this phylum[5], and the emergence of the monoderm cell envelope by multiple losses of the OM (Figure 5b). Moreover, they support the fact that LPS-OM have a very ancient origin in Bacteria and were inherited remarkably vertically throughout the diversification of the major diderm phyla.

Discussion

Past hypotheses have supported either a monoderm-first[1,3,4,28] or a diderm-first scenario[2] for Bacteria[5]. However, most of them did not consider the phylogenetic relationships among diderm and monoderm lineages and among outer envelope markers. The existence of three independent diderm lineages within the Firmicutes adds an important piece to the puzzle and shows that, at least in this phylum, the OM is ancestral and the monoderm phenotype is a derived character which arose multiple times independently through OM loss. The hypothesis of an ancestor with an LPS-OM is also supported by the evidence that the phylogeny of the Firmicutes is becoming increasingly populated by diderm lineages, notably at its deepest offshoots, and it cannot be excluded that even more will be discovered in the future.

How the OM would have been lost multiple times in the Firmicutes remains to be understood. Endosporulation was probably important in the transition between monoderms and diderms[3,29]. In fact, during the process of endosporulation the cell produces a spore that is transiently surrounded by a second membrane, which is then lost during maturation in sporulating monoderm Firmicutes, while is retained in sporulating diderm Firmicutes[30]. While previous hypotheses proposed that endosporulation would have allowed the emergence of the OM[3,4,29,30], we rather think that the opposite occurred, i.e. that viable accidents during endosporulation allowed multiple OM losses in the Firmicutes[5] (Figure 5b). The widespread presence and antiquity of endosporulation in this phylum (Supplementary Figure 1, Methods) could have allowed multiple occurrences of such accidents during its diversification, and this is probably the reason why the Firmicutes are currently the only phylum containing both monoderm and diderm envelopes (the presence of OMs in some Actinobacteria is likely an independent convergence, see below). The alternative scenario where the OM would have been acquired via multiple HGT events in the Firmicutes remains possible, but we believe it is less supported by our data. Moreover, rather than suggesting HGT, the clustering of the OM genes in diderm Firmicutes may indicate a tight coordination of the various OM biogenesis processes, which could represent a peculiarity of these lineages. Finally, if the OM was acquired multiple times by HGT, this is not simpler as a process to imagine, as it would have necessitated that all the complex machineries for OM biogenesis become immediately functional in a monoderm context, notably attaching the OM to the PG wall and stabilizing it, adapting existing flagella and secretion systems to span two membranes, and developing transport of key compounds to the OM.

It may be argued that the benefit of having an OM is such that there would have been no selective advantage in losing it. However, no advantage has to be invoked if the loss of the OM was the result of viable accidents making it unstable. These might have not led to a decrease in fitness so dramatic as to immediately disadvantage the resulting monoderm phenotype, in particular if accompanied (either preceded or followed) by an increase in thickness of the cell wall[2,5,10]. Moreover, the monoderm phenotype with a thick PG wall might have resulted advantageous in some specific environmental conditions (e.g. drought, high temperature). The absence of a core of protein families specific to monoderm Firmicutes may reflect the fact that different solutions were found independently to accommodate each of these multiple OM loss events, and it is interesting to note that PG structure is indeed highly variable in the Firmicutes[29], and likely also the arsenal of enzyme families needed to produce it and remodel it. Following loss of the OM, the genes involved in its biogenesis would have been progressively lost, and some perhaps repurposed for cell envelope functions in monoderm Firmicutes, a hypothesis that will need specific analysis and experimental evidence. The availability of genetic tools in the Negativicute *Veillonella parvula*[5] opens the way to test experimentally some of these hypotheses and will allow to gather precious insights about the biogenesis and functioning both the diderm and the monoderm cell envelope.

Whether the last bacterial common ancestor (LBCA) also had an OM has been unclear, notably due to uncertainties in the phylogenetic relationships among diderms and monoderms. We calculated a phylogeny including all the main bacterial phyla used in this study, and we inferred their diderm or monoderm nature of their cell envelope by mapping the presence or absence, respectively, of two key markers of the OM, BamA and LpxA (Figure 6a, Supplementary Table 9, Methods). Although the cell envelope remains uncharacterized in most phyla, notably the uncultured one for which the diderm or monoderm status can only be tentatively inferred in-silico[5,31], it is already clear that the presence of diderms is overwhelming in Bacteria as compared to monoderms. Whether the LBCA was already a diderm or a monoderm will require to firmly establish the root of Bacteria, a complex methodological issue that can only be solved by specific analyses. We chose here to display a root in between Terrabacteria and Gracilicutes, supported by our recent phylogenomic analysis[32]. Combined with our evidence that the LPS-OM did not spread across diderm bacteria by HGT but was rather inherited vertically (Figure 5a), this root may imply that the LBCA could have been a diderm (Figure 6b, top). Alternative roots have been proposed in the literature, such as a possible one lying in the monoderm Chloroflexi or in the Candidate Phyla Radiation[3,33,34], which could support an LBCA with one membrane (Figure 6b, bottom). Most importantly, and irrespective of whether the LBCA was a diderm or not, our results show that monoderm phyla do not constitute a natural group but are polyphyletic. This indicates that the monoderm cell envelope would in any case have appeared multiple times independently through OM loss, and working out all the evolutionary paths that led to these transitions is in our opinion the most important and challenging goal.

It remains unclear what was the mechanism involved in the loss of the OM in the monoderm phyla other than the Firmicutes, but it may have involved different types of viable accidents causing an instability of the OM. It is evident that the few known monoderm phyla are only present in the large clade of Terrabacteria which displays a larger range of cell envelopes with respect to the more homogenous Gracilicutes that are composed only of diderms mostly endowed with LPS (Figure 6a). This larger diversity possibly suggests that the Terrabacteria cell envelopes have retained some ancestral characteristics. As such, diderm Firmicutes could become good experimental models for the primordial bacterial cell envelope.

Currently, only the Firmicutes and the Actinobacteria display a mixture of monoderm and diderm lineages. However, the presence of the mycolic acid OM in Actinobacteria such as Corynebacteria[35] is likely an independent *de novo* origination, as these taxa do not possess any of the classical OM markers. The presence of complex cell envelope structure and potential OM-like structures has also been recently reported in the Chloroflexi[36], a phylum that lacks all classical OM markers and thought to be monoderm[37]. However, virtually nothing in known on the cell envelope structure of most bacterial phyla, in particular those with no representative cultured members, and obtaining ultrastructural data for these phyla is an important challenge of future research.

Finally, how the OM initially came into being remains unknown. It is possible that it arose from a simpler cell surrounded by a single membrane (e.g. monoderm type), but we have no means by using phylogeny to go this far back in time beyond the LBCA. However, we think it unlikely that endosporulation was at the origin of the OM in the LBCA[3,4], as today this type of sporulation that produces a transient OM is specific of the Firmicutes and likely originated in this phylum. Alternatively, early speculation by Blobel postulated that the very first cell was already surrounded by a double membrane through a mechanism involving the formation of a "gastruloid" vesicle and the fusion of its extremities[38]. So, it is possible that early life never went through a "simpler" monoderm phase but started already as diderm.

Some of these questions should be addressed in the future through a more systematic characterization of a wide range of cell envelopes, both monoderm and diderm, across all bacterial phyla -notably the uncultured ones-, combined with large-scale comparative genomics and phylogenetic analysis to fully reconstruct the evolutionary history that accompanied their evolution and the multiple transitions among them, as well as the development of experimental models from unexplored branches of the Tree of Life.

Acknowledgments

S.G., C.B. and J.W. wish to acknowledge funding from the French National Research Agency (ANR), project Fir-OM (ANR-16-CE12-0010) and from the Institut Pasteur Programmes Transversaux de Recherche (PTR 39-16). D.M. and D.P were supported by the Pasteur-Paris University (PPU) International PhD Program. This work used the computational and storage services (TARS cluster) provided by the IT department at Institut Pasteur, Paris.

Author Contributions

S.G. conceived the study. N.T. and D.M. carried out all comparative genomics and phylogeny analyses. J. W. helped with annotation of the OM markers. D.P and G.B. helped with genome reconstruction of two uncultured Limnochordia genomes in an earlier version of the study. P. A. assembled the DB Bacteria and calculated the reference tree shown in Figure 6. C.B. helped with annotation and overall supervision. N.T., C.B. and S.G. wrote the paper, with contribution from D.M and J.W. All authors have read and approved the manuscript.

Material and Methods

Updating the Firmicutes reference tree and identification of diderm lineages

We retrieved 1,639 genomes annotated as Firmicutes from the dataset of Parks et al.[13], deposited under NCBI BioProject PRJNA348753. These genomes were isolated from different environments and their genomes quality goes from partial to near complete (Supplementary Table 1). According to the taxonomy provided by Parks et al.[13], these genomes consist in 351 Bacilli, 980 Clostridia, 61 Erysipelotrichia, 1 Tissierellia, 62 Negativicutes, 1 Halanaerobiales and 183 annotated as unclassified Firmicutes. Because these UBA genomes were in the nucleotide format at the time of this analysis, we used Prodigal[39] with default parameters to predict genes. To analyze their phylogenetic placement within the reference phylogeny of Firmicutes, we added 230 complete genomes from representative of all families available in the NCBI databases as for December 2017, including 61 Bacilli, 86 Clostridia, 4 Tissierellia, 62 Negativicutes, 16 Halanaerobiales and the genome of the only available representative of Limnochordia, *Limnochorda pilosa*. This resulted in a databank of 1,869 genomes (DB LARGE Firmicutes) (Supporting Data).

Exhaustive Hidden Markov Model (HMM)-based homology searches (with an e-value cutoff of 1e-04) were carried out by using HMM profiles of the complete set of 54 bacterial ribosomal proteins from the Pfam 29.0 database[40] as queries using the HMMER package[41]. Absences were checked with TBLASTN[42]. 45 ribosomal proteins present in > 70% of the genomes were kept for analysis. 514 UBA Firmicutes genomes having less than 35 ribosomal proteins were considered as too partial and discarded from analysis. 13 taxa were included as outgroup (Supporting Data). The 45 ribosomal proteins of the 1,355 remaining UBA and outgroup genomes were aligned by CLUSTAL OMEGA[43] with default parameters and trimmed using BMGE-1.1[44] with the BLOSUM35 substitution matrix. The resulting trimmed alignments were concatenated into a supermatrix (1,368 taxa and 5,087 amino acid positions). A ML tree was generated using IQ-TREE v1.4.4[45], with the ultrafast bootstrap approximation[46] imposed by the very large size of the dataset and the C60-profile mixture model LG+C60+F+G, which is a variant of the CAT model[47] for ML analysis.

To identify new diderm lineages among the UBA genomes, we used HMMSEARCH with a cutoff e-value of 1e-04 to screen them for the proteins involved in the first conserved steps of LPS synthesis (LpxABCD) (TIGR01852, PF02684, PF03331, PF04613), and for other protein domains previously used as markers of Gram-negativity[7]: Omp85 (PF01103), POTRA (PF07244), ExbD (PF02472), Secretin (PF00263), TamB (PF04357) and TonB_C (PF03544).

In order to identify sporulating taxa, we used HMMSEARCH to screen the DB SMALL Firmicutes using the Pfam domain spo0A_C (PF08769) with the option -cut_ga, and we mapped the results onto the corresponding tree of Firmicutes (Supporting Data).

All trees were annotated using iToL[48].

Distribution of protein families and domains in diderm and monoderm Firmicutes

To carry out the large-scale comparative genomic analysis, we assembled a reduced databank of 316 genomes. It includes the 230 reference Firmicutes genomes and the newly identified UBA diderm Firmicutes (46 Limnochordia, 39 Negativicutes, 1 Halanaerobiales), for a total of 165 diderm and 151 monoderm taxa (DB SMALL Firmicutes) (Supporting Data). The 861,409 proteins contained in the DB SMALL Firmicutes were annotated by using eggNOG-Mapper[49] with default parameters. The eggNOG-Mapper tool uses precomputed orthologous groups and phylogenies from the eggNOG database[50] to transfer functional information from fine-grained orthologs only. In a second approach, Pfam domains were predicted for each protein using HMMSEARCH (e-value <= 1e-5) against the Pfam 29.0 database. The results of the two approaches were merged and each protein family was annotated according to the most frequent prediction of its members.

We performed all vs all pairwise comparisons of protein sequences contained in the DB SMALL Firmicutes using BLASTP v2.6.0 with default parameters. Protein families were assembled with SILIX v1.2.9[51]. Identity thresholds values from 30% to 60% with intervals of 5% were tested, with a coverage of at least 80%. The resulting protein families were then refined using HIFIX v1.0.5, which performs a three-step high-quality sequence clustering guided by network topology and multiple alignment likelihood[52]. To assess the most suitable identity threshold to group orthologous proteins, we tested different cutoffs by using as positive control the clustering of 16 ribosomal proteins commonly used in phylogenetic analyses and of the four core LPS proteins (LpxABCD). The identity threshold that maximized the number of true positives and minimized the number of false positives was 35%. This 80% coverage-35% identity cutoff cannot however completely exclude some false negatives or false positives. Applying this threshold resulted in 176,024 protein families.

From these, we retained the families present in at least five taxa, resulting in 15,964 families for further analysis (Supporting Data), and a presence/absence matrix was built. Among the 15,964 families, 41 were completely absent from monoderm Firmicutes while present in at least one member of all three diderm lineages ("strict core diderm families").

In order to relax this strict criterium we used two clustering approaches on the presence/absence matrix (HCLUST and K-MEANS, both implemented in R). HCLUST clusters hierarchically the families according to Jaccard distances calculated on the presence/absence matrix. We tested different cutoffs and chose the one that allowed gathering the four core LPS protein families (LpxABCD) in the same cluster as a positive control (number of generated clusters set to 3,500) (Supporting Data). Four clusters, HC5, HC7, HC8 and HC9 included at least one of the "strict core" diderm families with a large taxonomic distribution (present in more than 40% of the diderm Firmicutes genomes), totalling 120 families.

For the second method, we clustered families using K-MEANS (k = 500) based on Multiple Correspondence Analysis (MCA). As the K-MEANS approach involves defining a random set of starting points in a multidimensional space, 10 iterations were run. 173 protein families clustered with at least one of the 41 strict core diderm families and were present in more than 40% of the diderm Firmicutes genomes in all K-MEANS iterations. Among these families, 95 were common to HCLUST and K-MEANS.

In parallel to the distribution of protein families, we used a second approach based on protein domains. Because proteins can contain different domains or multiple occurrences of the same domain, we used three different approaches: in the first, we considered together all proteins containing exactly the same type and number of predicted domains (ALL); in the second, we considered together proteins containing the same domains even if some had more than one occurrence (COLLAPSED); in the third, we counted the same proteins as many times as the different domains they contain (SINGLE) (Supplementary Tables 7 and 8) (Supporting Data).

Finally, the four obtained datasets (protein families, and the three Pfam domain approaches) were used to identify the diderm and monoderm specific protein families and Pfam domains using the Pearson correlation coefficient (PCC) based on their presence/absence in the diderm and monoderm taxa (higher than 0.5). The conserved genomic locus for cell envelope components in the Firmicutes was assessed using GeneSpy[53].

Distribution of the OM cluster in Bacteria and evolutionary analysis

In order to study the distribution of the OM cluster, we built HMM profiles of all genes included in the OM cluster of diderm Firmicutes. Then, we used MacSyFinder [54] to identify clusters containing these genes in the DB SMALL Firmicutes and a second databank including 377 genomes representative of 58 main bacterial phyla (DB Bacteria) (Supporting Data). We defined a cluster as a system with at least three of these OM components with a separation no greater than five other genes.

Among the gene clusters identified above, we selected 11 OM markers most conserved across Bacteria (FabZ, KdsA, KdsB, KdsC, KdsD, LptB, LpxA, LpxB, LpxC, LpxD, LpxK_WaaA). For each of them, homologues were aligned with MAFFT using the L-INS-I option[55] and trimmed using BMGE-1.1. The resulting alignments were concatenated by allowing a maximum of six missing markers per taxon and leading to a supermatrix of 146 taxa and 1,998 amino acid position. A second matrix including the flagellar components FlgF, FlgG, FlgA, FlgH, FlgI and FlgJ was assembled by allowing a maximum of eight missing markers per taxon and included 122 taxa (because many do not have flagella) and 3,705 amino acid positions.

For the reference phylogeny of Bacteria, we used a concatenation of RNA polymerase subunits B, B', and IF-2 (2,144 amino acid positions and 377 taxa).

For all concatenations, ML trees were generated using IQ-TREE v1.4.4 with the profile mixture model LG+C60+F+G with ultrafast bootstrap supports calculated on 1,000 replicates from the original data.

In order to map cell envelope types onto the reference phylogeny Bacteria, we built an HMM profile for BamA using the sequences in the corresponding family described above, and used it together with the HMM profile of LpxA to screen the DB Bacteria with HMMSEARCH (e-value <= 1e-4). Results were then refined and absences checked with TBLASTN. All trees were annotated using iToL[56].

Declaration of Interests

The authors declare no competing interests.

Data availability

All raw data relative to this analysis (databanks, sequence accession numbers, sequence datasets and corresponding trees, protein families) are provided as Supporting Data.
https://data.mendeley.com/datasets/3pcn9779gc/draft?a=8cc3c448-b3e7-4b02-96fb-9b3a0af4625d

Figure legends

Figure 1: An updated reference phylogeny of the Firmicutes reveals a third diderm clade.

ML reference tree of the Firmicutes including 1,125 UBA Firmicutes genomes based on concatenation of 45 ribosomal proteins (1,368 taxa, 5,087 amino acid positions). The tree was inferred with IQ-TREE 1.4.4 using the LG+C60+F+G model. Node supports higher than 95% are displayed. The tree is rooted with representatives of major bacterial taxa. The scale bar corresponds to the average number of substitutions per site. Names in red correspond to groups containing UBA genomes only and no representative of any known family as for December 2017. Coloured triangles indicate the three diderm clades. For the corresponding full phylogeny see Supplementary Data.

Figure 2: Phyletic patterns of protein families highlight the functional core of the diderm Firmicutes OM.

(a) Venn diagram showing the distribution of 15,964 protein families of the Firmicutes pan-proteome in Negativicutes, Halanaerobiales, Limnochordia and monoderm Firmicutes.

(b) Excerpt of the HC-based distribution of the 15,964 protein families of the Firmicutes pan-proteome with a focus on the four HC clusters (HC5, HC7, HC8 and HC9) containing the 120 protein families composing the relaxed core (columns in red). (c) Distribution of the 15,964 protein families of the Firmicutes pan-proteome. Dots in purple and pink correspond to protein families with a PCC > 0.5 with the diderm or monoderm phenotype, respectively.

Figure 3: Distribution and functional annotation of the protein families and PFAM domains highly correlated with the diderm phenotype (a) and the monoderm phenotype (b).

(a) 52 protein families and Pfam domains correlated with the diderm phenotype (PCC > 0.5) and present in at least one member of each of the three diderm lineages. When a protein family and a Pfam domain corresponded, we chose to display the one with the best correlation. Presence in the OM gene cluster is also indicated. Superscripts indicate proteins which correspond to two different domains (TamB and OmpM).

(b) 51 protein families and Pfam correlated with the monoderm phenotype. When a protein family and a Pfam domain corresponded, the one with the best correlation is displayed.

Figure 4: A large OM gene cluster is a distinguished feature of all diderm Firmicutes.

Annotation of the genomic locus containing OM-related markers in all three diderm Firmicute lineages. Color codes: blue (LPS synthesis and transfer); orange (OMP biogenesis); red (OM tethering); green (Lipid asymmetry): pink (Flagellum); yellow (unclear function); white (no functional domains detected); gray (no direct link to OM); hashed: three conserved hypothetical proteins possibly related to the OM. Lateral bars mean separate loci in the genome. Vertical bars mean end of a contig in unclosed genomes. The organization of these genes in *E. coli* is given as a comparison. Monoderm Firmicutes (Clostridiales and Bacilli) are also displayed to illustrate the presence of the beginning of the OM cluster, a possible remnant of a diderm past. All accession numbers are provided in Supplementary Data.

Figure 5: Phylogenomic analysis does not support acquisition of the OM by horizontal gene transfer (a) and supports multiple and independent losses of the OM (b).

(a) ML tree based on the concatenation of 11 conserved OM markers (FabZ, KdsA, KdsB, KdsC, KdsD, LptB, LpxA, LpxB, LpxC, LpxD, LpxK_WaaA) including 14 diderm phyla and 1,998 amino acid positions. The tree was inferred with IQ-TREE 1.4.4 and the LG+C60+F+G model. Node supports higher than 70% are displayed. The scale bar corresponds to the average number of substitutions per site. In the absence of an outgroup, the tree is tentatively rooted in between Terrabacteria and Gracilicutes[32]. Even though the tree only contains diderm phyla, it shows that the presence of main OM markers predates the divergence of these phyla, including that of Firmicutes and that therefore an OM with LPS has an ancient origin. For the corresponding full phylogeny see Supplementary Data.

(b) Evolutionary scenario for the origin and evolution of the OM in the Firmicutes mapped on a schematic of the reference phylogeny in Figure 1. The ancestor of Firmicutes is indicated as a sporulating diderm with LPS. The LPS-OM was inherited in the three diderm lineages (Halanaerobiales, Limnochordia and Negativicutes), while it was lost three times independently to give rise to the classical monoderm cell envelope.

Figure 6: Distribution of monoderm and diderm cell envelopes across Bacteria and two potential evolutionary scenarios for their origin.

(a) ML reference tree of Bacteria, with cell envelope types mapped on it. The tree was obtained from a concatenation of RNA polymerase subunits B, B′, and translation initiation factor IF-2 (2,144 amino acids positions and 377 taxa). The tree was inferred with IQTREE and the LG+C60+G+F model. Node supports higher than 70% are displayed. The scale bar represents the average number of substitutions per site. The tree does not include an outgroup but is tentatively rooted between two large clades corresponding to Terrabacteria and Gracilicutes, as in Raymann et al.,[32]. For the corresponding full phylogeny see Supplementary Data. PVC s.l. (*sensu latu*) indicates the clade including the PVC superphylum (Planctomycetes, Verrucomicrobia, Chlamydiae) and relative phyla; FCB s.l. (*sensu latu*) indicates the clade including the FBC superphylum (Fibrobacteres, Bacteroidetes, Chlorobi) and relative phyla; Proteobacteria s.l. (*sensu latu*) indicates the clade including the Proteobacteria subdivisions and related phyla. For the corresponding full phylogeny see Supplementary Data.

(b) Evolutionary scenario mapped on a schematic version of the tree in (a). Different roots of the Bacteria have been proposed in the literature, we show here two alternative ones as an example: (i) in between Terrabacteria and Gracilicutes[32] (top) or (ii) in the branch leading to the CPR and Chloroflexi [33,34] (bottom). In the first scenario, the LBCA could have been a diderm. In the second scenario, the LBCA could have been a monoderm and the OM would have appeared just after the divergence of Chloroflexi and CPR. Note that both scenarios imply multiple losses of the OM. The second scenario also leaves open the possibility that the LBCA was a diderm and that the OM was lost in the branch leading to the Chloroflexi and the CPR.

References

1. Gupta, R. S. 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, International Journal of General and Molecular Microbiology* vol. 100 171–182 (2011).

2. Cavalier-Smith, T. The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int. J. Syst. Evol. Microbiol.* **52**, 7–76 (2002).

3. Tocheva, E. I., Ortega, D. R. & Jensen, G. J. Sporulation, bacterial cell envelopes and the origin of life. *Nat. Rev. Microbiol.* **14**, 535–542 (2016).

4. Errington, J. L-form bacteria, cell walls and the origins of life. *Open Biol.* **3**, 120143 (2013).

5. Megrian, D., Taib, N., Witwinowski, J., Beloin, C. & Gribaldo, S. One or two membranes? Diderm Firmicutes challenge the Gram-positive/Gram-negative divide. *Molecular Microbiology* vol. 113 659–671 (2020).

6. Mavromatis, K. *et al.* Genome analysis of the anaerobic thermohalophilic bacterium Halothermothrix orenii. *PLoS One* **4**, (2009).

7. Tocheva, E. I. *et al.* Peptidoglycan Remodeling and Conversion of an Inner Membrane into an Outer Membrane During Sporulation Elitza. *Cell* **146**, 799–812 (2012).

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

9. Helander, I. M., Hurme, R., Haikara, A. & Moran, A. P. Separation and characterization of two chemically distinct lipopolysaccharides in two Pectinatus species. *J. Bacteriol.* **174**, 3348–3354 (1992).

10. Antunes, L. C. *et al.* Phylogenomic analysis supports the ancestral presence of LPS-outer membranes in the firmicutes. *Elife* **5**, e14589 (2016).

11. Kojima, S. *et al.* 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. *J. Bacteriol.* **192**, 5953–5961 (2010).

12. Poppleton, D. I. *et al.* Outer membrane proteome of Veillonella parvula: A diderm firmicute of the human microbiome. *Front. Microbiol.* **8**, 1–17 (2017).

13. Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, (2017).

14. Yutin, N. & Galperin, M. Y. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ. Microbiol.* **15**, 2631–2641 (2013).

15. Watanabe, M., Kojima, H. & Fukui, M. 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 (2015).

16. Watanabe, M., Kojima, H. & Fukui, M. 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 (2016).

17. Silhavy, T. J., Kahne, D. & Walker, S. The bacterial cell envelope. *Cold Spring Harbor perspectives in biology* vol. 2 a000414–a000414 (2010).

18. Bos, M. P., Robert, V. & Tommassen, J. Biogenesis of the gram-negative bacterial outer membrane. *Annu. Rev. Microbiol.* **61**, 191–214 (2007).

19. Sperandeo, P., Martorana, A. M. & Polissi, A. The Lpt ABC transporter for lipopolysaccharide export to the cell surface. *Res. Microbiol.* (2019) doi:10.1016/j.resmic.2019.07.005.

20. Heinz, E., Selkrig, J., Belousoff, M. J. & Lithgow, T. Evolution of the translocation and assembly module (TAM). *Genome Biol. Evol.* **7**, 1628–1643 (2015).

21. Noinaj, N., Guillier, M., Barnard, T. J. & Buchanan, S. K. TonB-Dependent Transporters: Regulation, Structure, and Function. *Annu. Rev. Microbiol.* **64**, 43–60 (2010).

22. Hughes, G. W. *et al.* Evidence for phospholipid export from the bacterial inner membrane by the Mla ABC transport system. *Nat. Microbiol.* (2019) doi:10.1038/s41564-019-0481-y.

23. Mukherjee, S. & Kearns, D. B. The Structure and Regulation of Flagella in Bacillus subtilis . *Annu. Rev. Genet.* **48**, 319–340 (2014).

24. Jacquier, N., Yadav, A. K., Pillonel, T., Viollier, P. H. & Greub, G. A SpoIID Homolog Cleaves Glycan Strands at the Chlamydial Division Septum. *Mol. Biol. Physiol.* **10**, 1–16 (2019).

25. Delsuc, F., Brinkmann, H. & Philippe, H. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **6**, 361–375 (2005).

26. Cavalier-Smith, T. Rooting the tree of life by transition analyses. *Biol. Direct* **1**, 1–83 (2006).

27. Battistuzzi, F. U. & Hedges, S. B. A major clade of prokaryotes with ancient adaptations to life on land. *Mol. Biol. Evol.* **26**, 335–343 (2009).

28. Lake, J. A. Evidence for an early prokaryotic endosymbiosis. *Nature* **460**, 967–971 (2009).

29. Vollmer, W. Bacterial outer membrane evolution via sporulation? *Nat. Chem. Biol.* **8**, 14–18 (2012).

30. Tocheva, E. I. *et al.* Peptidoglycan remodeling and conversion of an inner membrane into an outer membrane during sporulation. *Cell* **146**, 799–812 (2011).

31. Sutcliffe, I. C. A phylum level perspective on bacterial cell envelope architecture. *Trends Microbiol.* **18**, 464–470 (2010).

32. Raymann, K., Brochier-Armanet, C. & Gribaldo, S. The two-domain tree of life is linked to a new root for the Archaea. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6670–6675 (2015).

33. Cavalier-Smith, T. & Chao, E. E. Y. Multidomain ribosomal protein trees and the planctobacterial origin of neomura (eukaryotes, archaebacteria). *Protoplasma* 621–753 (2020) doi:10.1007/s00709-019-01442-7.

34. Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1**, 1–6 (2016).

35. Vincent, A. T. *et al.* The mycobacterial cell envelope: A relict from the past or the result of recent evolution? *Front. Microbiol.* **9**, 1–9 (2018).

36. Gaisin, V. A., Kooger, R., Grouzdev, D. S., Gorlenko, V. M. & Pilhofer, M. Cryo-Electron Tomography Reveals the Complex Ultrastructural Organization of Multicellular Filamentous Chloroflexota (Chloroflexi) Bacteria. *Front. Microbiol.* **11**, 1–15 (2020).

37. Sutcliffe, I. C. Cell envelope architecture in the Chloroflexi: A shifting frontline in a phylogenetic turf war. *Environ. Microbiol.* **13**, 279–282 (2011).

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

39. Hyatt, D. *et al.* Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, (2010).

40. Finn, R. D. *et al.* The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2016).

41. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, (2010).

42. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

43. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, (2011).

44. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).

45. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

46. Minh, B. Q., Nguyen, M. A. T. & Von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).

47. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).

48. Letunic, I. & Bork, P. Interactive Tree of Life v2: Online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **39**, 1–4 (2011).

49. Huerta-Cepas, J. *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).

50. Huerta-Cepas, J. *et al.* EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293 (2016).

51. Miele, V., Penel, S. & Duret, L. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics* **12**, (2011).

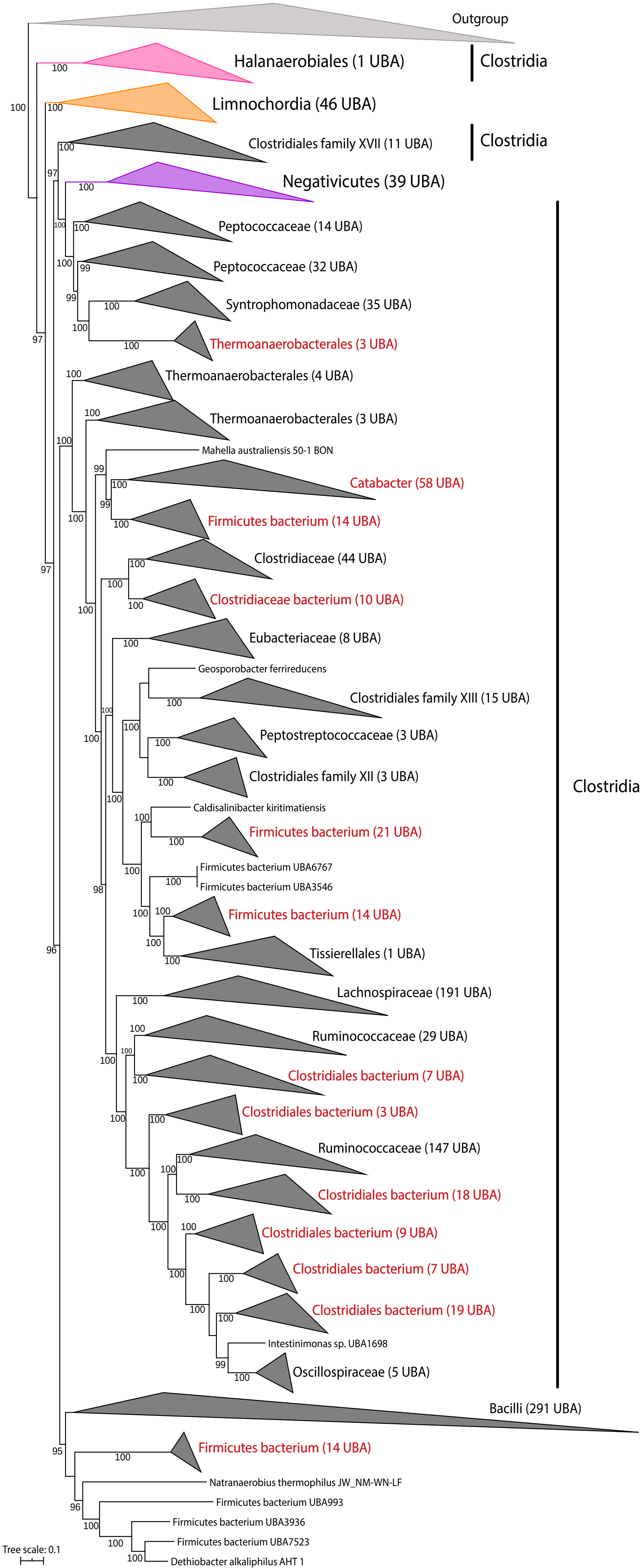
52. Miele, V. *et al.* High-quality sequence clustering guided by network topology and multiple alignment likelihood. *Bioinformatics* **28**, 1078–1085 (2012).

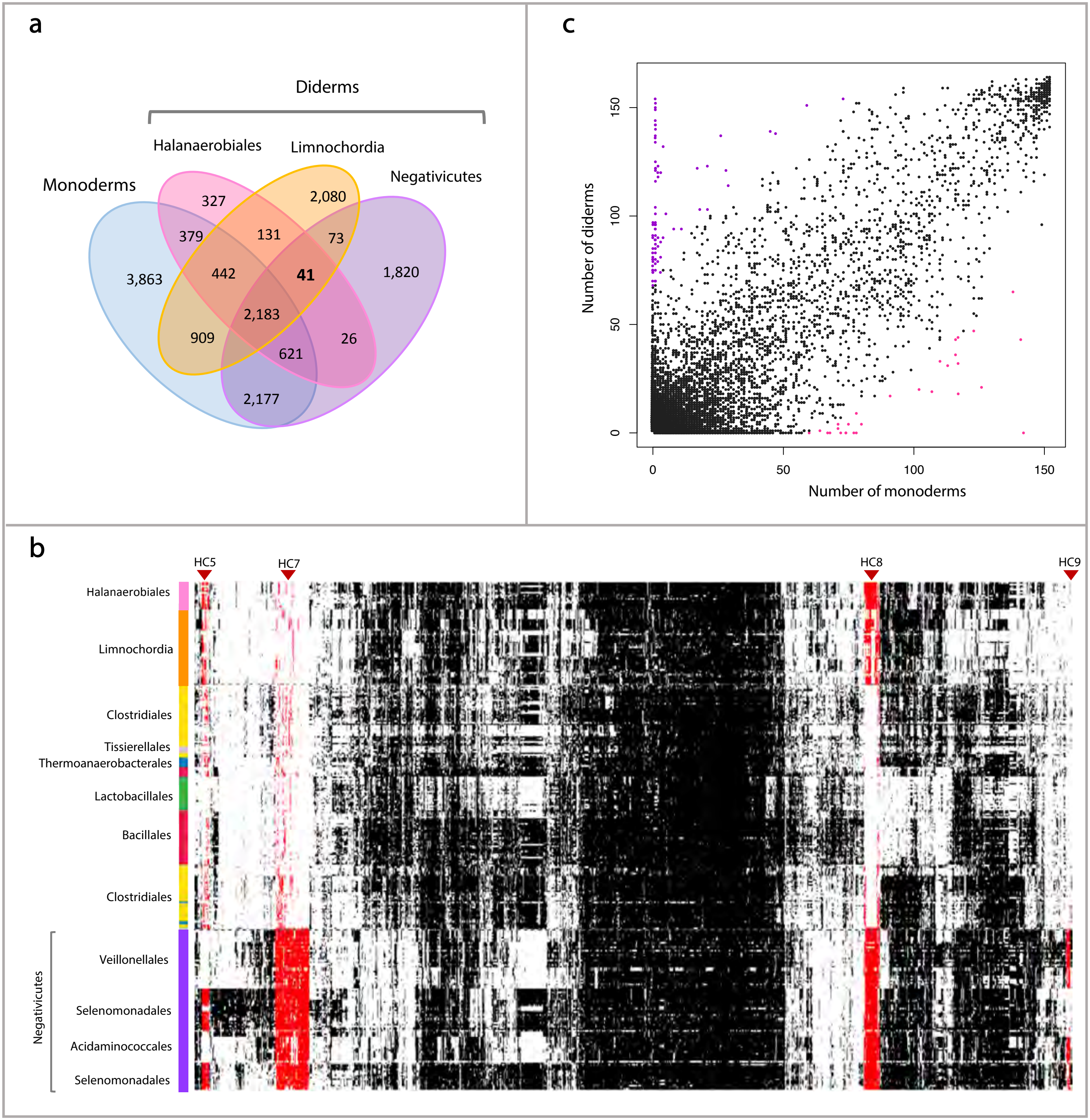
53. Garcia, P. S., Jauffrit, F., Grangeasse, C. & Brochier-Armanet, C. GeneSpy, a user-friendly and flexible genomic context visualizer. *Bioinformatics* **35**, 329–331 (2019).

54. Abby, S. S., Néron, B., Ménager, H., Touchon, M. & Rocha, E. P. C. MacSyFinder: A program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS One* **9**, (2014).

55. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

56. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).


Outgroup
Halanaerobiales (1 UBA)
Clostridia
Limnochordia (46 UBA)
Clostridiales family XVII (11 UBA)
Clostridia
Negativicutes (39 UBA)
Peptococcaceae (14 UBA)
Peptococcaceae (32 UBA)
Syntrophomonadaceae (35 UBA)
Thermoanaerobacterales (3 UBA)
Thermoanaerobacterales (4 UBA)
Thermoanaerobacterales (3 UBA)
Mahella australiensis 50-1 BON
Catabacter (58 UBA)
Firmicutes bacterium (14 UBA)
Clostridiaceae (44 UBA)
Clostridiaceae bacterium (10 UBA)
Eubacteriaceae (8 UBA)
Geosporobacter ferrireducens
Clostridiales family XIII (15 UBA)
Peptostreptococcaceae (3 UBA)
Clostridiales family XII (3 UBA)
Caldisalinibacter kiritimatiensis
Firmicutes bacterium (21 UBA)
Firmicutes bacterium UBA6767
Firmicutes bacterium UBA3546
Firmicutes bacterium (14 UBA)
Tissierellales (1 UBA)
Lachnospiraceae (191 UBA)
Ruminococcaceae (29 UBA)
Clostridiales bacterium (7 UBA)
Clostridiales bacterium (3 UBA)
Ruminococcaceae (147 UBA)
Clostridiales bacterium (18 UBA)
Clostridiales bacterium (9 UBA)
Clostridiales bacterium (7 UBA)
Clostridiales bacterium (19 UBA)
Intestinimonas sp. UBA1698
Oscillospiraceae (5 UBA)
Clostridia
Bacilli (291 UBA)
Firmicutes bacterium (14 UBA)
Natranaerobius thermophilus JW_NM-WN-LF
Firmicutes bacterium UBA993
Firmicutes bacterium UBA3936
Firmicutes bacterium UBA7523
Dethiobacter alkaliphilus AHT 1
Tree scale: 0.1

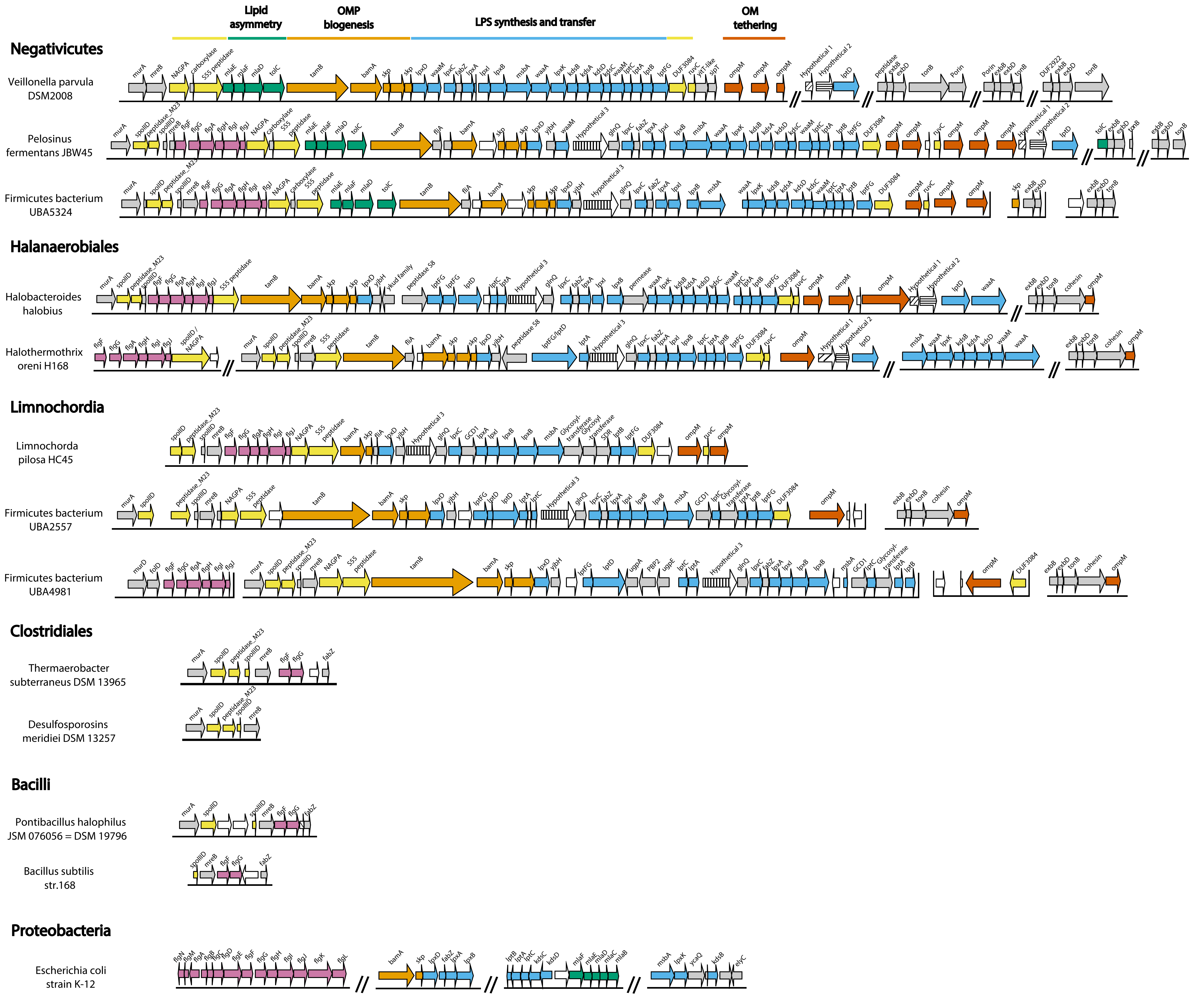

a
Diderms
Halanaerobiales
Limnochordia
Negativicutes
Monoderms
327
2,080
379
131
73
3,863
442
41
1,820
2,183
909
26
621
2,177
c
Number of diderms
Number of monoderms
0
50
100
150
0
50
100
150
b
HC5
HC7
HC8
HC9
Halanaerobiales
Limnochordia
Clostridiales
Tissierellales
Thermoanaerobacterales
Lactobacillales
Bacillales
Clostridiales
Negativicutes
Veillonellales
Selenomonadales
Acidaminococcales
Selenomonadales

a

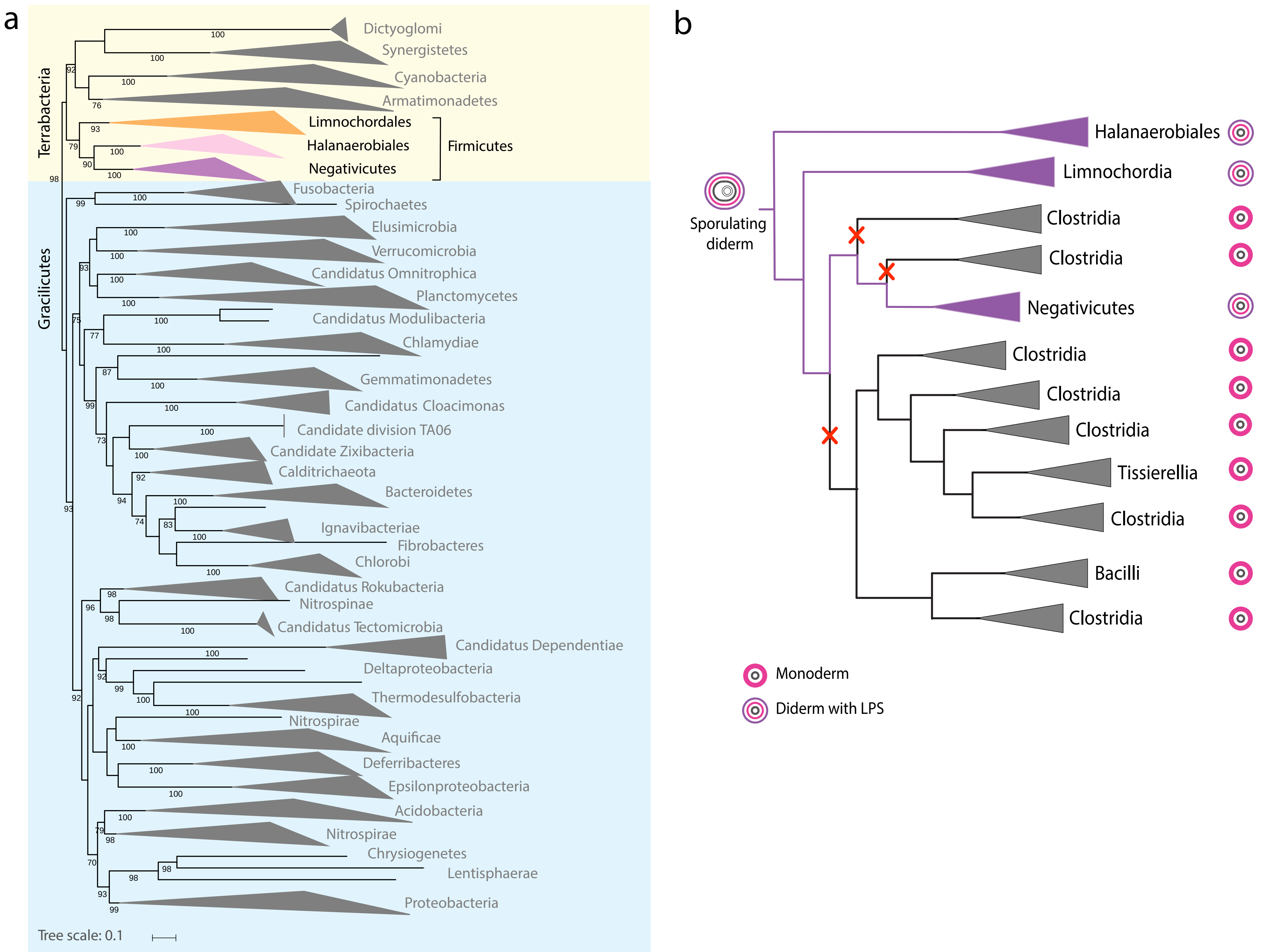
Protein Family/ Pfam Domain	Correlation coefficient	Diderm (165)	Monoderm (151)	Pfam Annotation	Putative Function	OM cluster
PF06835	0.97	159	0	LptC	LptC (LPS transport)	YES
PF01103	0.96	161	0	Bac_surface_Ag	BamA (OMP transport and insertion)	YES
PF11283	0.96	160	1	DUF3084	uncharacterized	YES
PF03968	0.92	155	2	OstA	LptA, LptD (LPS transport)	YES
FAM35001205_1	0.92	153	0	RuvX	RuvC (endonuclease)	YES
PF07660,PF03958,PF00263	0.92	157	3	STN,Secretin_N,Secretin	PilQ (Type II secretion)	
PF02684	0.91	152	2	LpxB	LpxB (LPS synthesis)	YES
FAM35003477_0	0.91	151	0	YjgP_YjgQ	LptFG (LPS transport)	YES
PF04357	0.91	151	0	TamB	TamB[1] (autotransporter assembly complex)	YES
FAM35003763_0	0.9	150	0	MotA_ExbB	ExbB (energy transducing Ton complex)	
PF03938	0.89	163	22	OmpH	Skp (OMP transport and insertion)	YES
PF02472	0.87	152	6	ExbD	ExbD (energy transducing Ton complex)	
FAM35001011_0	0.85	141	0	Peptidase_S55	Peptidase	YES
PF13502	0.85	141	2	AsmA_2	TamB[1] (autotransporter assembly complex)	YES
PF03544	0.84	140	1	TonB_C	TonB (energy transducing Ton complex)	
FAM35006918_0	0.83	137	0	Hexapep,Acetyltransf_11	LpxA (LPS synthesis)	YES
FAM35003933_0	0.83	138	0	Hexapep,LpxD	LpxD (LPS synthesis)	YES
FAM35004424_0	0.82	135	0	DUF1009	LpxI (LPS synthesis)	YES
PF03331	0.82	140	3	LpxC	LpxC (LPS synthesis)	YES
PF13505	0.79	131	1	OMP_b-brl	OM beta barrel	
FAM35002111_0	0.77	122	0	Lip_A_acyltrans	WaaM (LPS synthesis)	YES
FAM35000110_0	0.76	123	0	LpxK	LpxK (LPS synthesis)	YES
FAM35001087_0	0.76	123	0	TrbI	hypothetical 2 (Bacterial conjugation TrbI-like protein)	YES
FAM35002134_0	0.75	122	0	Hydrolase_3	KdsC (LPS synthesis)	YES
FAM35001758_0	0.75	124	1	CTP_transf_3	KdsB (LPS synthesis)	YES
PF04413	0.74	125	3	Glycos_transf_N	waaA (LPS synthesis)	YES
FAM35007211_0	0.74	121	1	DAHP_synth_1	KdsA (LPS synthesis)	YES
FAM35001331_0	0.73	119	1	Glyco_transf_9	glycosyl transferase	
FAM35000503_0	0.73	121	2	SIS,CBS	KdsD (LPS synthesis)	YES
PF05258	0.71	114	0	TonB_dep_Rec	TonB complex, OM receptor	
PF00593	0.71	126	8	DUF721	uncharacterized	
FAM35000108_2	0.66	138	25	Peptidase_M23	peptidase	YES
PF11741	0.65	145	34	AMIN	AmiC N-terminal domain	
PF11209	0.62	104	8	DUF2993	uncharacterized	
FAM35007523_0	0.61	123	21	RecX	modulates RecA activity	
FAM35012460	0.59	90	0	FlgH	FlgH (flagellum L-ring)	YES
FAM35002503	0.59	90	0	FlgI	FlgI (flagellum P-ring)	YES
PF00395	0.59	165	74	SLH	OmpM[2] (OM tethering)	YES
FAM35001613	0.58	88	0	Sigma70_r2,Sigma70_r4,Sigma70_r1_2	Transcription factor	
FAM35001477_0	0.56	152	58	Pribosyltran	Phosphoribosyl transferase domain	
FAM35002373_0	0.56	140	44	PS_pyruv_trans	Polysaccharide pyruvyl transferase	
PF02563,PF10531	0.55	78	0	Poly_export,SLBB	Polysaccharide biosynthesis/export protein, ligand binding	
FAM35000742_0	0.55	121	28	Germane	GerM (spore germination/cell division)	
PF04402	0.55	136	41	SIMPL	uncharacterized	
FAM35000495_0	0.54	139	46	DivIC	Septum formation	
PF13609	0.53	74	0	Porin_4	OmpM[2] (OM tethering)	YES
PF03797	0.52	73	1	Autotransporter	Type V(a) secretion	
FAM35006277	0.51	70	0	SurA_N_3	Outer membrane transporter	
PF03865	0.51	75	1	ShlB	Type V(b) secretion	
FAM35005930_0	0.51	155	72	B12-binding,Radical_SAM	unclear	
FAM35008032_0	0.51	115	28	Orn_Arg_deC_N	hypothetical 1 (Orn/Lys/Arg decarboxylase class-II family)	YES
FAM35004850_0	0.5	104	20	LysM	PG binding	

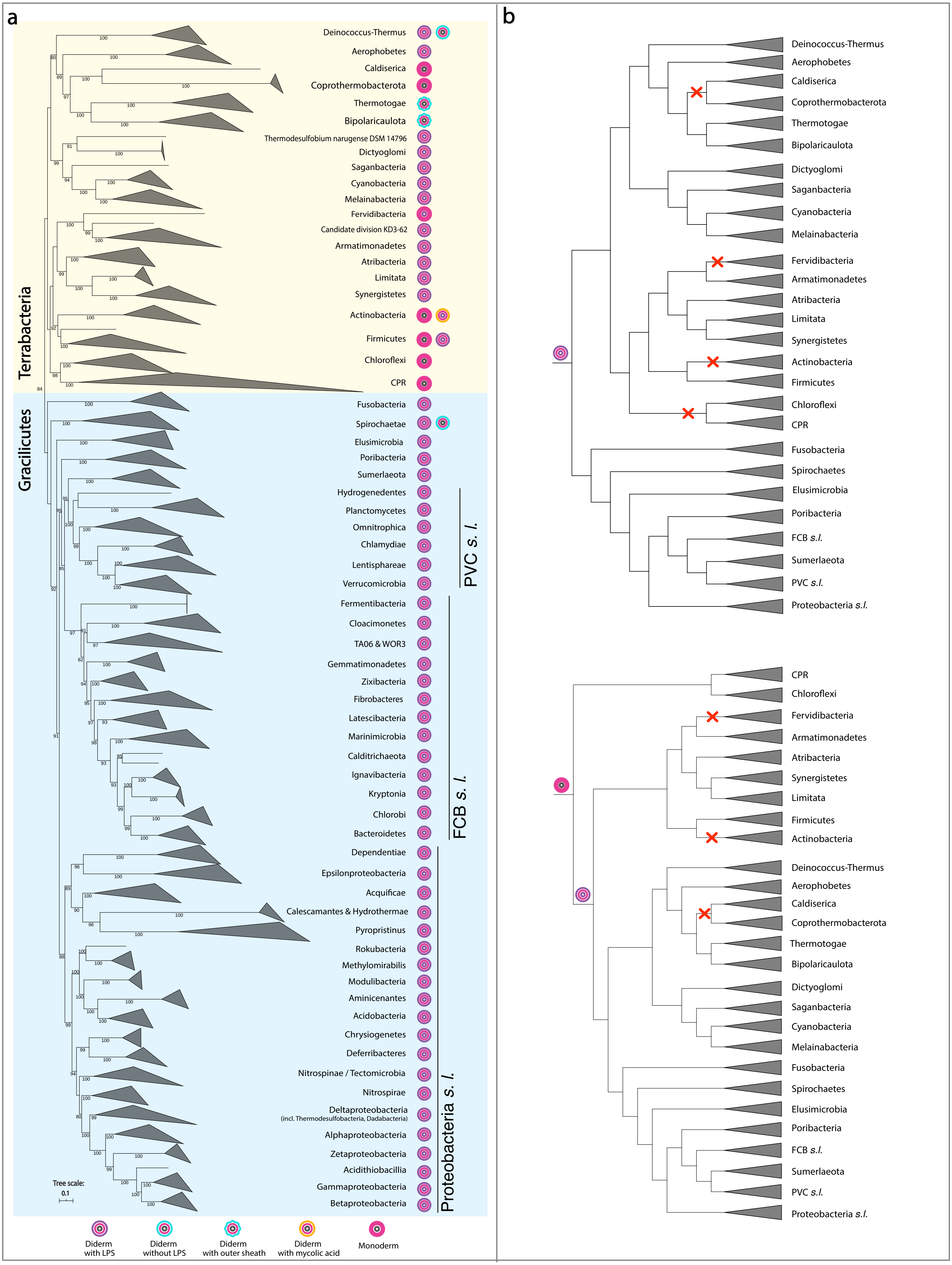
b

Protein Family/ Pfam Domain	Correlation coefficient	Diderm (165)	Monoderm (151)	Pfam Annotation
FAM35000114_0	0.94	0	142	S4 (RNA-binding, translation regulation)
PF04203.12	0.72	1	104	Sortase (attaches surface proteins to the cell wall)
FAM35004194_0	0.7	21	126	THUMP,UPF0020 (putative RNA methylase)
FAM35002871_0	0.67	44	140	CDP-OH_P_transf (CDP-alcohol phosphatidyltransferase)
FAM35000917_0	0.67	18	117	PLDc_N,PLDc_2 (phospholipase)
PF01170.17	0.66	36	133	UPF0020 (Putative RNA methylase)
PF08353.9	0.64	0	87	DUF1727 (C-terminus of bacterial proteins which include UDP-N-acetylmuramyl tripeptide synthase and the related Mur ligase)
PF07261.10	0.64	54	144	DnaB_2 (replication initiation and membrane attachment)
PF05389.11	0.64	0	86	MecA (negative regulator of competence)
PF03951.18	0.63	32	126	Gln-synt_N (glutamine synthetase)
FAM35005306_0	0.6	19	107	Hpr_kinase_N,Hpr_kinase_C (serine kinase)
PF00746.20	0.59	20	105	Gram_pos_anchor (LPXTG cell wall anchor motif)
FAM35000755	0.59	0	77	GntR,4HBT,CBS,DRTGG (unclear)
PF08245.11,PF08353.9	0.58	0	76	Mur_ligase_M,DUF1727 (Mur ligase middle domain)
PF01521.19	0.58	4	83	Fe-S_biosyn (iron-sulfur cluster biosynthesis)
PF02325.16	0.58	37	123	YGGT (repeat found in conserved hypothetical integral membrane proteins)
FAM35002460	0.57	0	74	GATase_3 (CobB/CobQ-like glutamine amidotransferase domain)
FAM35005063	0.57	4	80	Lactamase_B_2 (Beta-lactamase superfamily domain)
PF12679.6	0.57	45	126	ABC2_membrane_2 (ABC-2 family transporter protein)
FAM35001568	0.56	0	72	NAD_kinase (ATP-NAD kinase)
FAM35001032_0	0.56	21	101	Toprim,Topoisom_bac,Toprim_Crpt (topoisomerase)
PF01694.21	0.56	46	126	Rhomboid (intramembrane protease)
PF02517.15	0.56	76	148	Abi (abortive infection phage resistance)
PF06103.10	0.56	19	100	DUF948 (unknown function)
FAM35002806_1	0.56	31	113	PGI (phosphoglucose isomerase)
PF09648.9	0.55	3	75	YycI (regulate the essential YycFG two-component system in Bacillus subtilis)
PF04167.12	0.54	0	68	DUF402 (unknown function)
FAM35008547	0.54	0	68	Pribosyltran (Phosphoribosyl transferase domain)
FAM35004009_0	0.54	4	75	Radical_SAM,Radical_SAM_C (Radical_SAM)
FAM35003839	0.54	2	71	MazG (nucleotide pyrophosphohydrolase)
FAM35002605	0.54	0	67	Wzz,GNVR (O-antigen chain length, G-rich domain on putative tyrosine kinase)
PF07435.10	0.54	0	67	YycH (regulates the essential YycFG two-component system in Bacillus subtilis)
FAM35001229_1	0.53	65	138	PGM_PMM_I,PGM_PMM_II,PGM_PMM_III,PGM_PMM_IV
PF06207.10	0.53	4	73	DUF1002
PF07319.10	0.53	1	67	DnaI_N (Primosomal protein DnaI N-terminus)
PF01883.18	0.52	3	70	FeS_assembly_P (Iron-sulfur cluster assembly protein)
FAM35003271_0	0.52	47	123	TPK_catalytic,TPK_B1_binding Thiamin pyrophosphokinase)
FAM35005240_0	0.52	17	91	PNP_UDP_1 (Phosphorylase superfamily)
FAM35002295_0	0.52	4	71	Penicillinase_R (Penicillinase repressor)
PF01145.24	0.52	40	116	Band_7 (slipin or Stomatin-like integral membrane domain)
PF01643.16	0.52	2	67	Acyl-ACP_TE (Acyl-ACP thioesterase)
PF00355.25	0.52	36	74	Rieske (Rieske [2Fe-2S] domain)
PF03631.14	0.52	26	101	Virul_fac_BrkB (Virulence factor BrkB)
PF05975.11	0.51	7	75	EcsB (ABC transporter)
FAM35000288_1	0.51	1	64	Pyr_redox_2 (Pyridine nucleotide-disulphide oxidoreductase)
FAM35002436_1	0.51	9	78	AMP-binding,AMP-binding_C (AMP-binding enzyme)
PF06081.10	0.51	24	96	ArAE_1 (Aromatic acid exporter family member 1)
PF02687.20,PF02687.20	0.5	18	87	FtsX,FtsX (FtsX-like permease family)
FAM35006543	0.5	0	60	MMR_HSR1 (50S ribosome-binding GTPase)
PF04026.11	0.5	46	119	SpoVG (Stage V sporulation protein G)
PF15980.4	0.5	1	62	ComGF (Putative Competence protein ComGF)



Lipid asymmetry
OMP biogenesis
LPS synthesis and transfer
OM tethering
Negativicutes
Veillonella parvula DSM2008
murA
mreB
NAGPA
carboxylase
S55 peptidase
mlaE
mlaF
mlaD
tolC
tamB
bamA
skp
skp
lpxD
waaM
lpxC
fabZ
lpxA
lpxI
lpxB
msbA
waaA
lpxK
kdsB
kdsA
kdsD
kdsC
waaM
lptC
lptA
lptB
lptFG
DUF3084
ruvC
yitT-like
sipT
ompM
ompM
ompM
Hypothetical 1
Hypothetical 2
lptD
peptidase
exbB
exbD
tonB
Porin
Porin
exbB
exbD
tonB
DUF2922
exbB
exbD
tonB
Pelosinus fermentans JBW45
murA
spoIID
peptidase_M23
spoIID
mreB
flgF
flgG
flgA
flgH
flgI
NAGPA
carboxylase
S55 peptidase
mlaE
mlaF
mlaD
tolC
tamB
fliA
bamA
skp
skp
lpxD
yjbH
waaM
Hypothetical 3
glnQ
lpxC
fabZ
lpxA
lpxI
lpxB
msbA
waaA
lpxK
kdsB
kdsA
kdsD
kdsC
waaM
lptC
lptA
lptB
lptFG
DUF3084
ompM
ompM
ruvC
ompM
ompM
ompM
Hypothetical 1
Hypothetical 2
lptD
tolC
exbB
exbD
tonB
exbB
exbD
tonB
Firmicutes bacterium UBA5324
murA
spoIID
peptidase_M23
spoIID
mreB
flgF
flgG
flgA
flgH
flgI
flgJ
NAGPA
carboxylase
S55 peptidase
mlaE
mlaF
mlaD
tolC
tamB
fliA
bamA
skp
skp
lpxD
yjbH
Hypothetical 3
glnQ
lpxC
fabZ
lpxA
lpxI
lpxB
msbA
waaA
lpxK
kdsB
kdsA
kdsD
kdsC
waaM
lptC
lptA
lptB
lptFG
DUF3084
ompM
ruvC
ompM
ompM
skp
exbB
exbD
exbB
exbD
tonB
Halanaerobiales
Halobacteroides halobius
murA
spoIID
peptidase_M23
spoIID
flgF
flgG
flgA
flgH
flgI
S55 peptidase
tamB
bamA
skp
skp
lpxD
yjbH
ykud family
peptidase S8
lptFG
lptFG
lptD
lptC
lptA
Hypothetical 3
glnQ
lpxC
fabZ
lpxA
lpxI
lpxB
permease
waaA
lpxK
kdsB
kdsA
kdsD
kdsC
waaM
lptC
lptA
lptB
lptFG
DUF3084
ruvC
ompM
ompM
ompM
Hypothetical 1
Hypothetical 2
lptD
waaA
exbB
exbD
tonB
cohesin
ompM
Halothermothrix oreni H168
flgF
flgG
flgA
flgH
flgI
flgJ
spoIID / NAGPA
murA
spoIID
peptidase_M23
spoIID
mreB
S55 peptidase
tamB
fliA
bamA
skp
skp
lpxD
yjbH
peptidase S8
lptFG/lptD
lptA
Hypothetical 3
glnQ
lpxC
fabZ
lpxA
lpxI
lpxB
lptC
lptA
lptB
lptFG
DUF3084
ruvC
ompM
Hypothetical 1
Hypothetical 2
lptD
msbA
waaA
lpxK
kdsB
kdsA
kdsD
waaM
waaA
exbB
exbD
tonB
cohesin
ompM
Limnochordia
Limnochorda pilosa HC45
spoIID
peptidase_M23
spoIID
mreB
flgF
flgG
flgA
flgH
flgI
flgJ
NAGPA
S55 peptidase
bamA
skp
fliA
lpxD
yjbH
Hypothetical 3
glnQ
lpxC
GCD1
lpxA
lpxI
lpxB
lpxB
msbA
Glycosyl-transferase
Glycosyl-transferase
SDR
lptB
lptFG
DUF3084
ompM
ruvC
ompM
Firmicutes bacterium UBA2557
murA
spoIID
peptidase_M23
spoIID
mreB
NAGPA
S55 peptidase
tamB
bamA
skp
lpxD
yjbH
lptFG
lptD
lptD
lptA
lptC
Hypothetical 3
glnQ
lpxC
fabZ
lpxA
lpxI
lpxB
lpxB
msbA
GCD1
lptC
Glycosyl-transferase
lptA
lptB
lptFG
DUF3084
ompM
exbB
exbD
tonB
cohesin
ompM
Firmicutes bacterium UBA4981
murD
folD
flgF
flgG
flgA
flgH
flgI
flgJ
murA
spoIID
peptidase_M23
spoIID
mreB
NAGPA
S55 peptidase
tamB
bamA
skp
lpxD
yjbH
lptFG
lptD
ugpA
PBP2
ugpE
lptC
lptA
Hypothetical 3
glnQ
lpxC
fabZ
lpxA
lpxI
lpxB
lpxB
msbA
GCD1
lptC
Glycosyl-transferase
lptA
lptB
ompM
DUF3084
exbB
exbD
tonB
cohesin
ompM
Clostridiales
Thermaerobacter subterraneus DSM 13965
murA
spoIID
peptidase_M23
spoIID
mreB
flgF
flgG
fabZ
Desulfosporosins meridiei DSM 13257
murA
spoIID
peptidase_M23
spoIID
mreB
Bacilli
Pontibacillus halophilus JSM 076056 = DSM 19796
murA
spoIID
spoIID
mreB
flgF
flgG
fabZ
Bacillus subtilis str.168
spoIID
mreB
flgF
flgG
fabZ
Proteobacteria
Escherichia coli strain K-12
flgN
flgM
flgA
flgB
flgC
flgD
flgE
flgF
flgG
flgH
flgI
flgJ
flgK
flgL
bamA
skp
lpxD
fabZ
lpxA
lpxB
lptB
lptA
lptC
kdsC
kdsD
mlaF
mlaE
mlaD
mlaC
mlaB
msbA
lpxK
ycaQ
kdsB
elyC

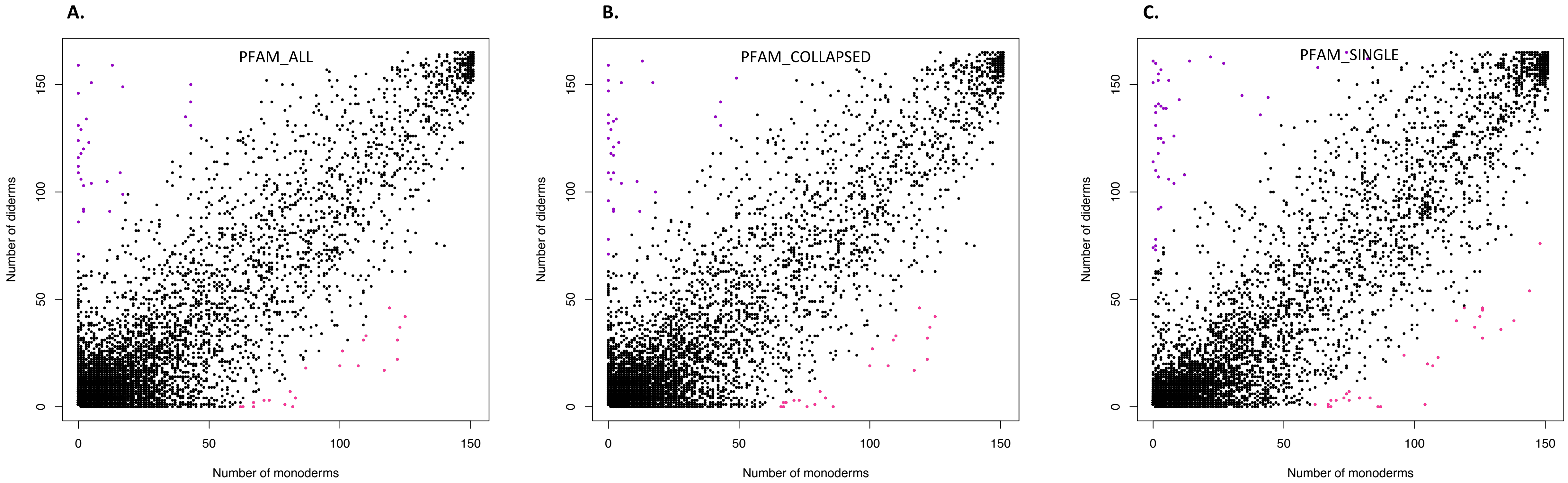

a
Terrabacteria
Gracilicutes
Dictyoglomi
100
Synergistetes
100
92
Cyanobacteria
100
Armatimonadetes
76
Limnochordales
93
Halanaerobiales
100
79
90
Negativicutes
100
Firmicutes
98
Fusobacteria
100
99
Spirochaetes
Elusimicrobia
100
Verrucomicrobia
100
93
Candidatus Omnitrophica
100
Planctomycetes
100
75
Candidatus Modulibacteria
100
77
Chlamydiae
100
87
Gemmatimonadetes
100
99
Candidatus Cloacimonas
100
Candidate division TA06
100
73
Candidate Zixibacteria
100
Calditrichaeota
92
94
Bacteroidetes
100
74
83
Ignavibacteriae
100
Fibrobacteres
Chlorobi
100
93
Candidatus Rokubacteria
98
Nitrospinae
96
98
Candidatus Tectomicrobia
100
Candidatus Dependentiae
100
Deltaproteobacteria
92
99
Thermodesulfobacteria
100
92
100
Nitrospirae
Aquificae
100
Deferribacteres
100
Epsilonproteobacteria
100
Acidobacteria
100
79
Nitrospirae
98
Chrysiogenetes
70
98
Lentisphaerae
98
93
Proteobacteria
99
Tree scale: 0.1
b
Sporulating diderm
Halanaerobiales
Limnochordia
Clostridia
Clostridia
Negativicutes
Clostridia
Clostridia
Clostridia
Tissierellia
Clostridia
Bacilli
Clostridia
Monoderm
Diderm with LPS


a
Terrabacteria
Deinococcus-Thermus
Aerophobetes
Caldiserica
Coprothermobacterota
Thermotogae
Bipolaricaulota
Thermodesulfobium narugense DSM 14796
Dictyoglomi
Saganbacteria
Cyanobacteria
Melainabacteria
Fervidibacteria
Candidate division KD3-62
Armatimonadetes
Atribacteria
Limitata
Synergistetes
Actinobacteria
Firmicutes
Chloroflexi
CPR
Gracilicutes
Fusobacteria
Spirochaetae
Elusimicrobia
Poribacteria
Sumerlaeota
Hydrogenedentes
Planctomycetes
Omnitrophica
Chlamydiae
Lentisphaereae
Verrucomicrobia
PVC s. l.
Fermentibacteria
Cloacimonetes
TA06 & WOR3
Gemmatimonadetes
Zixibacteria
Fibrobacteres
Latescibacteria
Marinimicrobia
Calditrichaeota
Ignavibacteria
Kryptonia
Chlorobi
Bacteroidetes
FCB s. l.
Dependentiae
Epsilonproteobacteria
Acquificae
Calescamantes & Hydrothermae
Pyropristinus
Rokubacteria
Methylomirabilis
Modulibacteria
Aminicenantes
Acidobacteria
Chrysiogenetes
Deferribacteres
Nitrospinae / Tectomicrobia
Nitrospirae
Deltaproteobacteria
(incl. Thermodesulfobacteria, Dadabacteria)
Alphaproteobacteria
Zetaproteobacteria
Acidithiobacillia
Gammaproteobacteria
Betaproteobacteria
Proteobacteria s. l.
Tree scale:
0.1
Diderm with LPS
Diderm without LPS
Diderm with outer sheath
Diderm with mycolic acid
Monoderm
b
Deinococcus-Thermus
Aerophobetes
Caldiserica
Coprothermobacterota
Thermotogae
Bipolaricaulota
Dictyoglomi
Saganbacteria
Cyanobacteria
Melainabacteria
Fervidibacteria
Armatimonadetes
Atribacteria
Limitata
Synergistetes
Actinobacteria
Firmicutes
Chloroflexi
CPR
Fusobacteria
Spirochaetes
Elusimicrobia
Poribacteria
FCB s.l.
Sumerlaeota
PVC s.l.
Proteobacteria s.l.
CPR
Chloroflexi
Fervidibacteria
Armatimonadetes
Atribacteria
Synergistetes
Limitata
Firmicutes
Actinobacteria
Deinococcus-Thermus
Aerophobetes
Caldiserica
Coprothermobacterota
Thermotogae
Bipolaricaulota
Dictyoglomi
Saganbacteria
Cyanobacteria
Melainabacteria
Fusobacteria
Spirochaetes
Elusimicrobia
Poribacteria
FCB s.l.
Sumerlaeota
PVC s.l.
Proteobacteria s.l.

Tree scale: 0.1
spo0A
Acetohalobium arabaticum DSM 5501
Halobacteroidaceae bacterium T328-2
Selenihalanaerobacter shriftii
Halobacteroides halobius DSM 5150
Halonatronum saccharophilum DSM 13868
Orenia marismortui DSM 5156
Orenia metallireducens
Halothermothrix sp. UBA3952
Halothermothrix orenii H 168
Halarsenatibacter silvermanii
Halanaerobium hydrogeniformans
Halanaerobium salsuginis
Halanaerobium praevalens DSM 2228
Halanaerobium saccharolyticum subsp. saccharolyicum DSM6643
Halanaerobium kushneri
Halanaerobium congolense
Halanaerobium sp. T82-1
Firmicutes bacterium UBA5301
Firmicutes bacterium UBA3928
Firmicutes bacterium UBA4881
Firmicutes bacterium UBA5447
Firmicutes bacterium UBA4953
Firmicutes bacterium UBA5499
Firmicutes bacterium UBA5473
Firmicutes bacterium UBA5237
Limnochorda pilosa
Firmicutes bacterium UBA3918
Firmicutes bacterium UBA3899
Firmicutes bacterium UBA3914
Firmicutes bacterium UBA4996
Firmicutes bacterium UBA5420
Firmicutes bacterium UBA5474
Firmicutes bacterium UBA5493
Firmicutes bacterium UBA5322
Firmicutes bacterium UBA4891
Firmicutes bacterium UBA5292
Firmicutes bacterium UBA4939
Firmicutes bacterium UBA4882
Firmicutes bacterium UBA4862
Firmicutes bacterium UBA3943
Firmicutes bacterium UBA5486
Firmicutes bacterium UBA5293
Firmicutes bacterium UBA7641
Firmicutes bacterium UBA4976
Firmicutes bacterium UBA5268
Firmicutes bacterium UBA5388
Firmicutes bacterium UBA5435
Firmicutes bacterium UBA3919
Firmicutes bacterium UBA6811
Firmicutes bacterium UBA3915
Firmicutes bacterium UBA2557
Firmicutes bacterium UBA2560
Firmicutes bacterium UBA4981
Firmicutes bacterium UBA4971
Firmicutes bacterium UBA5449
Firmicutes bacterium UBA5516
Firmicutes bacterium UBA3956
Firmicutes bacterium UBA1393
Firmicutes bacterium UBA5275
Firmicutes bacterium UBA4980
Firmicutes bacterium UBA5389
Firmicutes bacterium UBA5362
Firmicutes bacterium UBA5253
Firmicutes bacterium UBA5289
Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446
Kyrpidia tusciae DSM 2912
Effusibacillus pohliae DSM 22757
Tumebacillus flagellatus
Thermobacillus composti KWC4
Saccharibacillus kuerlensis DSM 22868
Paenibacillus polymyxa SC2
Thermicanus aegyptius DSM 12793
Brevibacillus brevis NBRC 100599
Aneurinibacillus aneurinilyticus ATCC 12856
Shimazuella kribbensis DSM 45090
Exiguobacterium antarcticum B7
Tuberibacillus calidus DSM 17572
Sporolactobacillus laevolacticus DSM 442
Anaerobacillus macyae
Marinococcus halotolerans DSM 16375
Salsuginibacillus kocurii DSM 18087
Amphibacillus xylanus NBRC 15112
Lentibacillus jeotgali
Oceanobacillus iheyensis HTE831
Ornithinibacillus scapharcae TW25
Terribacillus aidingensis
Sediminibacillus halophilus NSP9.3
Thalassobacillus devorans MSP14
Pontibacillus halophilus JSM 076056 = DSM 19796
Salimicrobium jeotgali
Bacillus subtilis subsp. subtilis str. 168
Caldibacillus debilis DSM 16016
Geobacillus stearothermophilus 10
Jeotgalibacillus alimentarius
Planococcus antarcticus DSM 14505
Sporosarcina newyorkensis 2681
Kurthia huakuii LAM0618
Solibacillus silvestris StLB046
Ureibacillus thermosphaericus str. Thermo-BF
Gemella bergeriae ATCC 700627
Jeotgalicoccus marinus DSM 19772
Staphylococcus aureus subsp. aureus NCTC 8325
Macrococcus caseolyticus JCSC5402
Listeria monocytogenes serotype 4b str. F2365
Atopococcus tabaci DSM 17538
Allofustis seminis DSM 15817
Alloiococcus otitis ATCC 51267
Dolosigranulum pigrum ATCC 51524
Granulicatella adiacens ATCC 49175
Bavariicoccus seileri DSM 19936
Aerococcus urinae ACS-120-V-Col10a
Abiotrophia defectiva ATCC 49176
Facklamia languida CCUG 37842
Eremococcus coleocola DSM 15696
Carnobacterium maltaromaticum LMA28
Catellicoccus marimammalium M35_04_3
Enterococcus faecium DO
Lactococcus garvieae Lg2
Streptococcus pneumoniae R6
Lactobacillus casei ATCC 334
Pediococcus claussenii ATCC BAA-344
Weissella ceti
Oenococcus kitaharae DSM 17330
Fructobacillus fructosus KCTC 3544
Leuconostoc carnosum JB16
Natranaerobius thermophilus JW_NM-WN-LF
Dethiobacter alkaliphilus AHT 1
Tepidanaerobacter acetatoxydans Re1
Mahella australiensis 50-1 BON
Caldicellulosiruptor obsidiansis OB47
Thermoanaerobacterium thermosaccharolyticum DSM 571
Caldanaerobacter subterraneus subsp. tengcongensis MB4
Thermoanaerobacter brockii subsp. finnii Ako-1
Ruminiclostridium thermocellum AD2
[Clostridium] cellulolyticum H10
Ethanoligenens harbinense YUAN-3
Oscillibacter valericigenes Sjm18-20
Faecalibacterium prausnitzii L2-6
Ruminococcus bromii L2-63
[Eubacterium] siraeum 70/3
Clostridium lentocellum DSM 5427
Lachnoclostridium phytofermentans ISDg
[Eubacterium] eligens ATCC 27750
[Clostridium] saccharolyticum WM1
[Eubacterium] rectale DSM 17629
Butyrivibrio proteoclasticus B316
[Eubacterium] cellulosolvens 6
Blautia sp. YL58
Blautia obeum A2-162
Candidatus Arthromitus sp. SFB-mouse-Japan
Clostridium novyi NT
Clostridium baratii str. Sullivan
Clostridium perfringens SM101
Clostridium pasteurianum BC1
Clostridium acetobutylicum EA 2018
Clostridium botulinum A2 str. Kyoto
Clostridium kluyveri NBRC 12016
Eubacterium limosum KIST612
Acetobacterium woodii DSM 1030
Caldisalinibacter kiritimatiensis
Ezakiella massiliensis
Murdochiella sp. Marseille-P2341
Finegoldia magna ATCC 53516
Parvimonas micra
Peptoclostridium acidaminophilum DSM 3953
Clostridioides difficile ATCC 43255
Acetoanaerobium sticklandii
Filifactor alocis ATCC 35896
Geosporobacter ferrireducens
Eubacterium nodatum ATCC 33099
Eubacterium sulci ATCC 35585
Eubacterium infirmum F0142
Eubacterium brachy ATCC 33089
Eubacterium saphenum ATCC 49989
Mogibacterium sp. CM50
Mogibacterium timidum ATCC 33093
Fusibacter sp. 3D3
Acidaminobacter hydrogenoformans DSM 2784
Symbiobacterium thermophilum IAM 14863
Sulfobacillus acidophilus DSM 10332
Sulfobacillus thermosulfidooxidans DSM 9293
Thermaerobacter marianensis DSM 12885
Thermaerobacter subterraneus DSM 13965
Firmicutes bacterium UBA5324
Anaerosporomusa subterranea
Acetonema longum DSM 6540
Sporomusaceae bacterium UBA864
Sporomusaceae bacterium UBA1887
Sporomusaceae bacterium UBA5989
Sporomusaceae bacterium UBA7701
Sporolituus thermophilus DSM 23256
Thermosinus carboxydivorans Nor1
Sporomusa sphaeroides DSM 2875
Sporomusa ovata DSM 2662
Sporomusa acidovorans
Dendrosporobacter quercicolus
Propionispora vibrioides
Pelosinus fermentans JBW45
Pelosinus fermentans UBA6563
Pelosinus fermentans DSM 17108
Succinispira mobilis DSM 6222
Acidaminococcus intestini DSM 21505
Acidaminococcus sp. UBA7736
Acidaminococcus sp. UBA2733
Acidaminococcus timonensis
Acidaminococcus fermentans UBA1700
Acidaminococcus fermentans DSM 20731
Acidaminococcus fermentans UBA7757
Phascolarctobacterium succinatutens YIT 12067
Phascolarctobacterium sp. CAG:207
Phascolarctobacterium sp. CAG:266
Succiniclasticum ruminis DSM 9236
Succiniclasticum ruminis
Acidaminococcaceae bacterium UBA7746
Acidaminococcaceae bacterium UBA2812
Acidaminococcaceae bacterium UBA2873
Acidaminococcaceae bacterium UBA2842
Megamonas rupellensis DSM 19944
Megamonas hypermegale UBA1814
Megamonas hypermegale DSM 1672
Propionispira raffinosivorans DSM 20765
Megamonas sp. UBA1792
Megamonas sp. UBA1796
Megamonas sp. UBA1799
Schwartzia succinivorans DSM 10502
Anaerovibrio lipolyticus LB2005
Selenomonas sputigena ATCC 35185
Selenomonadales bacterium UBA7018
Selenomonadales bacterium UBA3815
Selenomonas artemidis F0399
Selenomonas flueggei ATCC 43531
Selenomonas sp. oral taxon 136
Selenomonas noxia F0398
Centipeda periodontii DSM 2778
Selenomonas sp. oral taxon 920
Selenomonas sp. oral taxon 138 str. F0429
Selenomonas ruminantium AB3002
Selenomonas ruminantium subsp. lactilytica TAM6421
Selenomonas ruminantium UBA3715
Selenomonadaceae bacterium UBA2897
Selenomonas bovis DSM 23594
Selenomonadaceae bacterium UBA7260
Selenomonadaceae bacterium UBA7286
Mitsuokella sp. oral taxon 131 str. W9106
Mitsuokella multacida DSM 20544
Mitsuokella sp. UBA4253
Mitsuokella sp. UBA4225
Negativicoccus massiliensis
Negativicoccus succinicivorans DORA_17_25
Veillonellaceae bacterium UBA5809
Allisonella histaminiformans
Dialister pneumosintes
Dialister micraerophilus DSM 19965
Veillonellaceae bacterium DNF00626
Dialister invisus DSM 15470
Dialister sp. UBA7259
Dialister sp. UBA7295
Dialister sp. UBA1703
Dialister sp. UBA2909
Dialister succinatiphilus YIT 11850
Veillonella magna DSM 19857
Veillonella seminalis ACS-216-V-Col6b
Veillonella sp. oral taxon 780 str. F0422
Veillonella montpellierensis DSM 17217
Veillonella sp. oral taxon 158 str. F0412
Veillonella parvula DSM 2008
Veillonella tobetsuensis
Veillonella atypica
Megasphaera paucivorans
Megasphaera sp. BV3C16-1
Megasphaera micronuciformis F0359
Megasphaera sp. UBA4382
Megasphaera sp. UBA4264
Megasphaera sp. UBA4233
Megasphaera genomosp. type_1 str. 28L
Megasphaera cerevisiae DSM 20462
Megasphaera sp. UBA1696
Megasphaera sp. MJR8396C
Megasphaera elsdenii UBA3716
Megasphaera elsdenii DSM 20460
uncultured Megasphaera sp.
Megasphaera sp. UBA4708
Megasphaera sp. UBA5897
Megasphaera sp. UBA6440
Thermincola potens JR
Thermincola ferriacetica
Carboxydothermus hydrogenoformans Z-2901
Candidatus Desulforudis audaxviator MP104C
Ammonifex degensii KC4
Desulfotomaculum nigrificans CO-1-SRB
Desulfotomaculum aeronauticum DSM 10349
Desulfotomaculum hydrothermale Lam5 = DSM 18033
Desulfotomaculum acetoxidans DSM 771
Pelotomaculum thermopropionicum SI
Desulfotomaculum arcticum DSM 17038
Desulfotomaculum geothermicum DSM 3669
Peptococcaceae bacterium BRH_c4b
Desulfotomaculum thermosubterraneum DSM 16057
Desulfotomaculum copahuensis
Syntrophomonas wolfei subsp. wolfei str. Goettingen G311
Syntrophothermus lipocalidus DSM 12680
Desulfitibacter sp. BRH_c19
Moorella thermoacetica ATCC 39073
Heliobacterium modesticaldum Ice1
Dehalobacter sp. TeCB1
Syntrophobotulus glycolicus DSM 8271
Peptococcaceae bacterium CEB3
Desulfitobacterium metallireducens DSM 15288
Desulfitobacterium hafniense DCB-2
Desulfitobacterium dichloroeliminans LMG P-21439
Desulfosporosinus acidiphilus SJ4
Desulfosporosinus sp. OL
Desulfosporosinus sp. OT
Desulfosporosinus sp. I2
Desulfosporosinus lacus DSM 15449
Desulfosporosinus meridiei DSM 13257

Supplementary figure 1: Phylogeny of the Firmicutes based on a reduced taxon sampling corresponding to the genomes used for the construction of the protein families. Related to Figure 1.

Maximum likelihood tree based on concatenation of 45 ribosomal proteins (329 taxa, 5,382 amino acid characters), inferred with IQ-TREE 1.4.4 using the LG+C60+F+G model. Dots at nodes represent bootstrap values (BV) higher than 80% calculated on 100 replicates of the original dataset. The scale bar corresponds to the average number of substitutions per site. Spo0A was mapped on this tree by scanning the taxa for the presence of Spo0A_C (PF08769.11) using HMMSEARCH. Red bars indicate the presence of spo0A.


A.
PFAM_ALL
B.
PFAM_COLLAPSED
C.
PFAM_SINGLE
Number of diderms
0
50
100
150
Number of monoderms
0
50
100
150

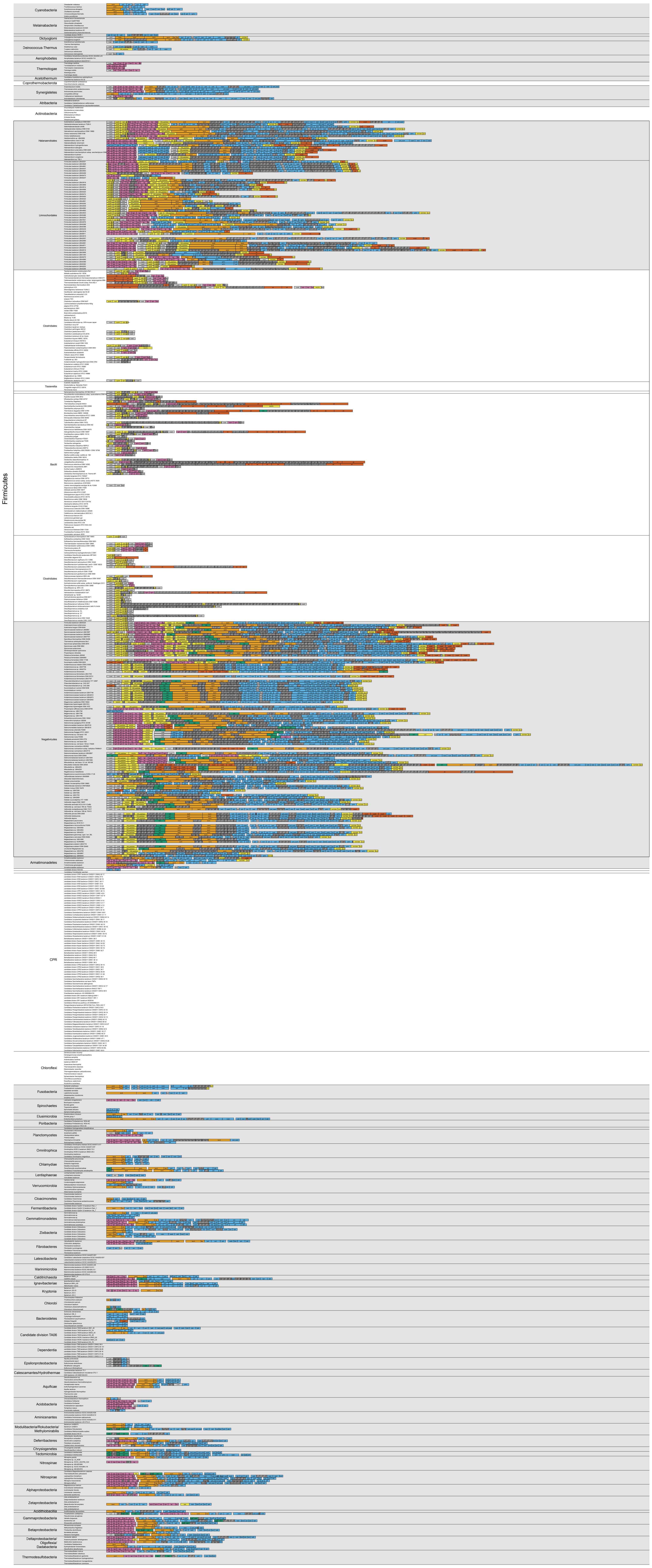
Supplementary figure 2: Distribution of PFAM domains according to their presence in 165 diderm and 151 monoderm Firmicutes. Related to Figure 3.

Dots in purple and pink correspond to domains with a correlation higher than 0.5 with the presence or absence of an OM respectively. The three approaches are represented: PFAM_ALL, PFAM_COLLAPSED and PFAM_SINGLE.

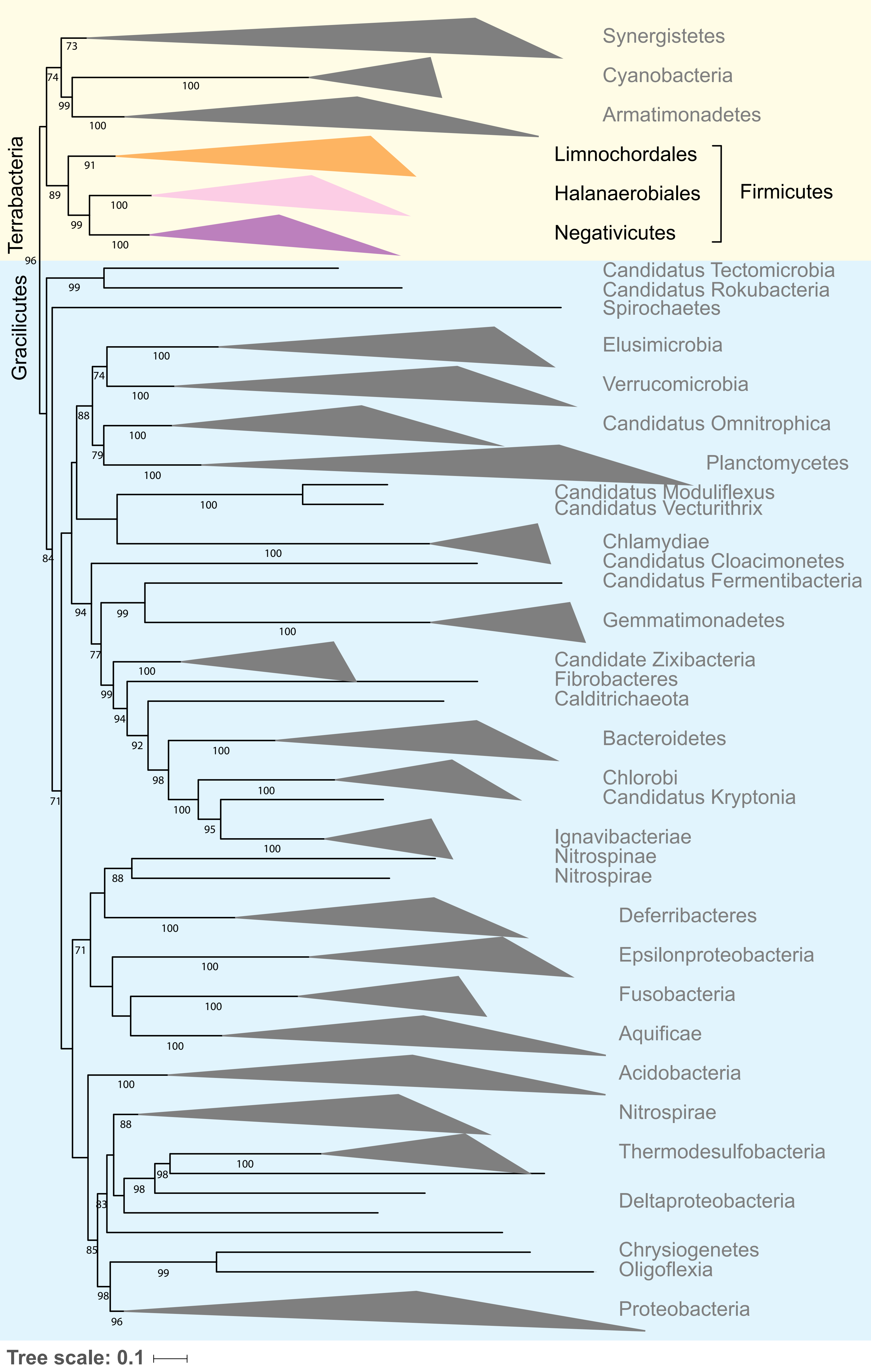
(a) Distribution of 56,513 PFAM_ALL domains.

(b) Distribution of 51,413 PFAM_COLLAPSED domains.

(c) Distribution of 14,808 PFAM_SINGLE domains.


Cyanobacteria
Melainabacteria
Dictyoglomi
Deinococcus-Thermus
Aerophobetes
Thermotogae
Acetothermum
Coprothermobacterota
Synergistetes
Atribacteria
Actinobacteria
Firmicutes
Halanaerobiales
Limnochordales
Clostridiales
Tissierella
Bacilli
Clostridiales
Negativicutes
Armatimonadetes
CPR
Chloroflexi
Fusobacteria
Spirochaetes
Elusimicrobia
Poribacteria
Planctomycetes
Omnitrophica
Chlamydiae
Lentisphaerae
Verrucomicrobia
Cloacimonetes
Fermentibacteria
Gemmatimonadetes
Zixibacteria
Fibrobacteres
Latescibacteria
Marinimicrobia
Calditrichaeota
Ignavibacteriae
Kryptonia
Chlorobi
Bacteroidetes
Candidate division TA06
Dependentia
Epsilonproteobacteria
Calescamantes/Hydrothermae
Aquificae
Acidobacteria
Aminicenantes
Modulibacteria/Rokubacteria/
Methylomirabilis
Deferribacteres
Chrysiogenetes
Tectomicrobia
Nitrospinae
Nitrospirae
Alphaproteobacteria
Zetaproteobacteria
Acidithiobacillia
Gammaproteobacteria
Betaproteobacteria
Deltaproteobacteria/
Oligoflexia/
Dadabacteria
Thermodesulfobacteria

Supplementary figure 3: Organization of OM clusters in Bacteria. Related to Figure 4. OM clusters are listed for all taxa in DB BACTERIA and DB SMALL Firmicutes. Phyla in gray are diderm, those in white are monoderm. Color codes are as in Figure 5.


Terrabacteria
Gracilicutes
73
Synergistetes
74
100
Cyanobacteria
99
100
Armatimonadetes
91
Limnochordales
89
100
Halanaerobiales
Firmicutes
99
100
Negativicutes
96
Candidatus Tectomicrobia
99
Candidatus Rokubacteria
Spirochaetes
100
Elusimicrobia
74
100
Verrucomicrobia
88
100
Candidatus Omnitrophica
79
100
Planctomycetes
Candidatus Moduliflexus
100
Candidatus Vecturithrix
Chlamydiae
100
84
Candidatus Cloacimonetes
Candidatus Fermentibacteria
94
99
100
Gemmatimonadetes
77
Candidate Zixibacteria
100
Fibrobacteres
99
Calditrichaeota
94
92
Bacteroidetes
100
98
Chlorobi
100
Candidatus Kryptonia
100
71
95
Ignavibacteriae
100
Nitrospinae
88
Nitrospirae
Deferribacteres
100
71
Epsilonproteobacteria
100
Fusobacteria
100
Aquificae
100
Acidobacteria
100
Nitrospirae
88
Thermodesulfobacteria
100
98
98
83
Deltaproteobacteria
85
Chrysiogenetes
99
Oligoflexia
98
Proteobacteria
96
Tree scale: 0.1

Supplementary figure 4: Phylogeny of OM cluster components including flagellar proteins. Related to Figure 5.

Maximum likelihood tree based on concatenation of 17 proteins (FabZ, KdsA, KdsB, KdsC, KdsD, LptB, LpxA, LpxB, LpxC, LpxD, LpxK, FlgF, FlgG, FlgA, FlgH, FlgI and FlgJ) comprising 122 taxa and 3,705 amino acid positions, inferred with IQ-TREE 1.4.4 using the model LG+C60+F+G and ultrafast bootstrap (UFB) supports calculated on 1000 replicates of the original dataset. UFB higher than 70% are displayed. The scale bar corresponds to the average number of substitutions per site.