Origin and novel transport pathways of bacterial hopanoids

Evan C. Lawrence[1,2], Huiqiao Pan[1], Noah Ollikainen[3], and Brittany J. Belin[1,2*]

[1]Department of Embryology, Carnegie Institution for Science, Baltimore, MD 21218
[2]Department of Biology, Johns Hopkins University, Baltimore, MD 21218
[3]Laboratory of Molecular Biology and Immunology, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224

* corresponding author: belin@carnegiescience.edu

ABSTRACT

Hopanoids are bacterial lipids that fortify membranes and enhance stress resistance. Derivatives of hopanoids known as "geo-hopanes" are abundant signatures of ancient bacteria in sediments, yet there are conflicting views on whether they are markers of specific taxa or environments. Here we analyze conservation of hopanoid biosynthesis across bacterial genomes using modern taxonomic tools. We find that hopanoids most likely originated in an ancestor of the pseudomonadota, and that tolerance of high osmolarity is the most common feature of hopanoid-producing strains. Additionally, microsynteny with hopanoid-related loci revealed new hopanoid-associated gene families, which we term *hpnT* and *hpnS*. Structural predictions and the restriction of these gene families to diderms suggests they participate in hopanoid trafficking, potentially forming a pathway analogous to Mla proteins in *E. coli*.

INTRODUCTION

Hopanoids are steroid-like isoprenoid lipids in diverse bacteria that mediate membrane biophysical processes, including resistance to environmental stress[1,2]. In sediments, ancient hopanoid derivatives (the "geo-hopanes") are abundant, with the oldest biomarkers dating to >1.6 Ga[3]. Their longevity in geologic time is due to their highly stable, pentacyclic hydrocarbon core, which is generated by consolidation and cyclization of six 5-carbon isoprene units. The simplest hopanoids contain only this 30-carbon core, sometimes with minor modifications (methylation or C-C bond desaturation), and are known as the "short" or C_{30} class. Hopanoids with ribose-derived side chains form the "extended" or C_{35} class. The C_{35} hopanoids are structurally diverse, including the addition of hydroxyl or amine moieties or covalent attachment to other biomolecules.

Many studies have examined the extent to which hopanoid biomarkers can be used to infer the identities of ancient microbes and their environments[4]. Such inferences rely on understanding which bacteria produce which hopanoids, with the assumption that the distribution of biological processes across phyla – and their compatibilities with certain environments (*e.g.*, oxic vs. anoxic habitats) – are conserved throughout time. Early *in vivo* studies hypothesized that geo-hopanes are markers of the advent of oxygenic photosynthesis. Specifically, hopanoids methylated at the C2 position (2Me-hopanoids) have been used as markers for marine cyanobacteria[3], and the proliferation of 2Me-hopanoid derivatives during oceanic anoxic events (OAEs) led to the suggestion that marine cyanobacteria expanded in these periods[5].

Increased availability of bacterial genomes and metagenomes, and the discovery of hopanoid biosynthesis enzymes (**Table 1**), have broadened our understanding of where and by which species hopanoids can be produced[6,7]. We now know that hopanoid-producing organisms are taxonomically diverse, making it difficult to classify geo-hopanes as markers of specific metabolisms or environmental contexts. Whether hopanoids can be used as markers of marine cyanobacteria remains a subject of debate. Several omics-based

studies argue against use of hopanoids as cyanobacterial markers, suggesting that 2-Me hopanoids originated in the alphaproteobacteria[8] and that hopanoids are more generally markers for nitrogen cycling[7,9-12]. Other, conflicting studies have reinforced the connection between hopanoids and cyanobacteria[13]. Most recently it was suggested that 2-Me hopanoids still can be used as cyanobacterial markers prior to 750 Ma[14], based on observations that (i) the evolutionary relationship between cyanobacterial hopanoid C2-methylases (*hpnP*) closely matches that of cyanobacteria genomes, indicating that *hpnP* was present in the last common cyanobacterial ancestor, and (ii) *hpnP*-encoding proteobacteria evolved much later than *hpnP*-encoding cyanobacteria[15].

A thorough understanding of the distribution and biological contexts of hopanoid production is necessary to understand the ancient ecosystems in which hopanoids were deposited. However, bacterial phylogenetic relationships have been drastically revised since the publication of studies on this topic[16]. This update reflects a shift from 16S rRNA-based classification to whole genome analysis in bacterial taxonomics, and it necessitates both new analysis of hopanoid distribution and re-evaluation of prior work. Here we analyze the distribution of hopanoid biosynthesis enzymes in fully sequenced bacteria. Phylogenetics of four hopanoid-related proteins suggest that the capacity to produce hopanoids likely originated in an ancestor of the pseudomonadota, and that tolerance of high osmolarity environments, nitrogen cycling, and single-carbon metabolism are common to hopanoid-producing species. Additionally, by evaluating microsynteny between known hopanoid-related genes and uncharacterized bacterial genes, we identify two new hopanoid-associated families, which we term *hpnT* and *hpnS*. Structural predictions and the restriction of these gene families to diderms suggests they participate in hopanoid trafficking, potentially forming a pathway analogous to Mla proteins in *E. coli* and other species.

RESULTS

Hopanoid biosynthesis genes are distributed across diverse taxa of sequenced bacteria

To determine the distribution of hopanoid biosynthesis genes, we collected high quality, fully sequenced bacterial genomes from the Joint Genome Institute's Integrated Microbial Genomes (IMG) database[17](**Table S1**). Due to prolific inconsistencies between taxonomic classification databases, we developed a hybrid genome classification approach based on two methods: whole genome classification with the Genome Taxonomy Database Toolkit (GTDB-Tk)[18], which assigns taxa through the genus rank based on similarly across 120 highly conserved bacterial genes, and species-level classifications via custom 16S rRNA analysis. Classification through the species rank was necessary because of the extreme bias towards model species in sequencing data: nearly 40% of bacterial genomes in IMG represent model strains from only six genera of human pathogens (Staphylococcus, Mycobacterium, Escherichia, Salmonella, Streptococcus, and Pseudomonas).

Our hybrid classification method yielded 5,568 unique species groups (**Table S1**) from the initial set of 64,376 genomes. Within each species group, we calculated the percentage of genomes with at least one gene annotated with a hopanoid-associated protein domain or pathway. These domain conservation percentages were re-calculated at the family level or higher, such that percentages reflect conservation across 5 or more genomes wherever possible (**Extended Data Table 1**). IMG annotates genomes with four unique protein domains associated with squalene-hopene cyclase (SHC), the enzyme that synthesizes C_{30} hopanoids. Though all four annotations are found in similar genomes in most phyla, the pfam and TIGRFAM IDs for SHC (**Table 1**) have an expanded range within the bacillota and fusobacteriota (**Fig. S1**). We speculated that this was due to mis-annotation of squalene cyclases involved in sporulene biosynthesis (SqhC proteins) and/or sterol biosynthesis enzymes, which are both closely related to SHC. To assess this interpretation, we constructed a phylogenetic tree of all proteins annotated as SHC in IMG (**Extended Data Fig. 1**). We found that most of such proteins from bacillota formed a separate clade, whereas only a smaller subset of bacillota proteins - primarily from the

family Alicyclobacillaceae - are in the main SHC branch. Given that *Alicyclobacillus acidocaldarius* is the only experimentally verified hopanoid-producer in the bacillota[19,20], we believe that only this smaller subset of proteins contains true SHC enzymes.

By this metric, we calculated that ~8% (449/5568) of species within our dataset have the genetic capacity to produce hopanoids (**Fig. 1**). SHC-containing species are most abundant within the pseudomonadota, with high conservation across alphaproteobacteria and within the gammaproteobacterial clade formerly classified as betaproteobacteria (primarily the *Burkholderia* and *Paraburkholderia* spp.). SHC is also highly conserved within the desulfobacteriota (formerly the deltaproteobacteria and thermodesulfobacteriota), acidobacteriota, cyanobacteriota, and planctomycetota. Lower levels of SHC conservation are found in former PVP group phyla (planctomycetota, verrucomicrobiota, omnitrophota, and poribacteria) and bacteroidota, whereas SHC is absent from campylobacteriota, spirochaetota, synergistota, thermotogota, and deinococcota.

For C_{35} hopanoid-associated domains (HpnGHIJK), there is generally more agreement in the annotation distribution. The HpnG, HpnI, and HpnJ domains, all of which are involved in downstream modification of C_{35} hopanoids, were only identified in genomes that contain the SHC and HpnH domains required for the first steps of C_{35} hopanoid production. HpnK also is involved in downstream modification of C_{35} hopanoids, but its distribution includes some likely spurious annotations within bacillota. We infer from these distributions that C_{35} hopanoid production and downstream modification is most common in the pseudomonadota, desulfobacterota, acidobacteriota, and cyanobacteriota. Some actinomycetota appear to have the capacity to produce ribosyhopane via HpnH but lack the enzymes for downstream modification.

Of the other verified hopanoid-associated domains included in IMG, those involved in hopanoid transport across the inner membrane of diderm bacteria (HpnN) or hopanoid methylation (HpnP and HpnR) have the sparsest taxon distribution. HpnN proteins are common in the pseudomonadota and desulfobacterota. HpnP and HpnR domains can be found in pseudomonadota and desulfobacterota, whereas only HpnP domains are found in cyanobacteriota and only HpnR domains are found in actinomycetota.

Hopanoid production likely arose in pseudomonadota and in high osmolarity environments

Our genome analysis suggests that hopanoid production is most conserved, and that hopanoid structures are most diverse, within the pseudomonadota. We therefore hypothesized that hopanoid biosynthesis originated within this phylum and was then laterally transferred to other species. To address this hypothesis, we collected sequences for SHC, HpnH, HpnN, HpnP, and HpnR domain proteins. We aligned sequences for each protein and generated phylogenetic trees. Maximum likelihood trees for all five protein families suggest that the oldest family members indeed belong to the alpha- or gammaproteobacteria (**Fig. 2**), a result supported by Bayesian analysis of SHC sequences (**Extended Data Fig. 2**).

To identify lateral transfer events, we compared our protein trees to the GTDB bacterial species tree (bac120) generated by alignment of 120 conserved gene regions[16,21](**Extended Data Fig. 3**). Given that the protein tree topologies clearly do not match the phylum relationships, we infer that hopanoid biosynthesis genes underwent a history of lateral transfer, with multiple gene transfers occurring within pseudomonadota before the genes were transferred to other phyla. An exception to this trend is HpnN, which includes proteins that form two distinct clades: a major clade dominated by the pseudomonadota, and a smaller clade containing desulfobacterota and uncharacterized species, potentially indicating that HpnN arose independently in these phyla. Though SHC and HpnH trees also include two clades, both clades originate in the pseudomonadota.

To find shared features of hopanoid-producing bacteria, we identified all protein domain families that are significantly enriched within genomes encoding SHC and other hopanoid biosynthesis genes (**Table S2**). This analysis revealed several protein functional categories that are positively correlated with hopanoid production (**Table 2**). The most enriched domains in SHC-containing genomes were associated with synthesis and modification of hopanoids, as well as of squalene, isoprenoid precursors (farnesyl and geranyl

diphosphate), and other cyclic isoprenoids such as carotenoids or sterols. An enrichment of domains associated with production of B vitamins and S-adenosyl-methionine (SAM) metabolism also was apparent, likely relating to the role of vitamin B_{12} and SAM as co-factors for hopanoid biosynthesis proteins HpnH, HpnP, and HpnR.

The second most significant category of correlated proteins includes those involved in osmotic pressure sensing and response. Specifically, domains involved in biosynthesis of the osmolyte trehalose and other saccharides involved in osmotic stress (*e.g.*, osmoregulated periplasmic glucans) are enriched in hopanoid-producers to a similar level as other isoprenoid biosynthesis domains. This category also includes the osmotic stress-regulated two-component system KdpPE, which controls potassium flux across the membrane via the potassium pump KdpFABC, and various transporters involved in divalent cation homeostasis. Interestingly, osmo-responsive protein domains are correlated with SHC across nearly all hopanoid-producing taxa.

Specific bacterial metabolisms were also highlighted by our analysis. Domains involved in utilization of single-carbon substrates and in nitrogen cycling are common to SHC-encoding genomes, and the biggest group of SHC-correlated metabolic domains are involved in aerobic respiration, such as the high efficiency terminal oxidases required for nitrogen fixation. To test this relationship between SHC and the nitrogen cycle, we measured the growth of our recently validated hopanoid-deficient strain of *Bradyrhizobium diazoefficiens*[22,23] on several nitrogen cycle intermediates (**Fig. S2**). We find that hopanoid-deficient mutants grow slowly in the presence of nitrite, but have no growth defect in the presence of ammonia or nitrate. *B. diazoefficiens* was unable to grow in hydroxylamine under any condition. Our results indicate that hopanoids may be linked to nitrite metabolism, supporting other recent work[10-12].

Microsynteny suggests novel hopanoid transport factors in diderm bacteria

Roughly 16% of the top 1000 domains positively correlated with SHC were annotated as uncharacterized or domains of unknown function (DUFs) (**Table S2**); collectively, this group was the most enriched category in SHC-containing genomes. We hypothesized that any DUFs directly involved in hopanoid biosynthesis would be adjacent to other hopanoid-associated genes within the genome. To identify such potential DUFs, we performed a targeted analysis of protein domains found near to the *shc* gene locus. We collected domain annotations for all genes within 20 ORFs up- or down-stream of *shc* in each genome, categorizing these genes into families based on shared domain annotations (**Table S3**).

Our analysis revealed four uncharacterized gene families that are commonly found near the *shc* locus and have potentially hopanoid-related functions (**Extended Data Fig. 4**). Two of these families are annotated with domains that already were computationally predicted to be related to hopanoids. The *hpnL* gene family includes genes annotated with the uncharacterized HpnL domain in TIGRFAM (TIGR03476); this domain does not exist in other protein family or pathway databases, and in InterPro is instead annotated as a member of the Lysylphosphatidylglycerol synthetase/glycosyltransferase AglD (IPR022791) domain family. The AglD InterPro domain includes proteins involved in addition of lysine addition to the phospholipid phosphatidylglycerol and in microbial S-layer modification. The *hpnM* gene family is annotated with the InterPro Hopanoid biosynthesis-associated membrane protein HpnM domain (IPR017842), which appears to be a subset of Toluene tolerance Ttg2/phospholipid-binding protein MlaC (IPR008869) domain proteins. Two new candidate hopanoid gene families, which we name here *hpnS* and *hpnT*, have not previously been associated with hopanoids. The *hpnS* family genes share the domain annotation MlaA lipoprotein (IPR007428). The *hpnT* genes are annotated with Domain of unknown function DUF2147 (IPR019223), and crystal structure of the DUF2147 domain protein HP1028 from *Helicobacter pylori*[24] suggests these domains belong to the lipocalin family, a group of proteins that transport hydrophobic small molecules[25].

In *E. coli* and other diderms, MlaC and MlaA are components of an intermembrane phospholipid transport system[26,27], whereas lipocalins can serve as extracellular lipid carriers in eukaryotes[28]. We thus hypothesized that HpnM, HpnS, and HpnT may form components of inner-to-outer membrane transport

systems for hopanoids, in concert with the inner membrane hopanoid transporter HpnN. This possibility is supported by the restriction of these proteins to diderms and the presence of predicted signal peptides, indicating a likely periplasmic localization (**Fig. S3**). Our subsequent analysis of the conservation of *hpnM*, *hpnS*, and *hpnT* genes within the vicinity of *hpnN* loci suggests that they evolved in distinct taxa (**Fig. 3**): the lipocalin-like HpnT protein is conserved in the rhizobiales order of alphaproteobacteria, whereas both HpnM and HpnS co-occur in the sphingomonadales, rhodospirillales, and azospirillales orders of alphaproteobacteria and across gammaproteobacteria.

MlaFEDB proteins form a complex[29] that moves phospholipids through the inner membrane, delivered by the periplasmic shuttle protein MlaC, which in turn ferries phospholipids from the MlaA lipoprotein that associates with outer membrane proteins[27,30,31] (**Fig. 4A**). We speculated that HpnM and HpnS might also form a shuttle and OM anchor protein system, which would require them to bind one another at the OM (**Fig 4B**). To test this hypothesis, we used AlphaFold 3[32], which predicted a high probability that the *Burkholderia multivorans* HpnM protein binds both the hopanoid diploptene and HpnS in a manner similar to the predicted MlaA/MlaC dimer[33] (**Fig. 4C**). *Bradyrhizobium diazoefficiens* HpnT is predicted to bind diploptene and may also act as a periplasmic shuttle, though it has no obvious cognate OM binding partner. These structural predictions support the hypothesis that HpnM, HpnS, and HpnT are novel hopanoid trafficking factors.

MATERIALS & METHODS

Bacterial genome selection

Genomes were selected for analysis by filtering the Joint Genome Institute's Integrated Microbial Genomes (IMG) database (https://img.jgi.doe.gov/)[17] for high quality bacterial assemblies with at least 1000 genes. Genome statistics were downloaded and genome accessions were further filtered to remove accessions where <75% of genes were associated with one or more TIGRFAM or Pfam annotation(s). Full genome sequences for the remaining accessions (totaling 64,376 genomes) were downloaded from IMG along with their associated ORF, protein domain, and functional pathway annotations.

Genome classification

Full genome sequences were annotated using the GTDB Toolkit v2.4.0+ (gtdb-tk, https://github.com/Ecogenomics/GTDBTk)[18] to determine taxonomic classifications at the phylum, class, order, and genus ranks. For species classification, 16S rRNA similarity was used. All 16S rRNA gene sequences were exported from the 64,376 genome set. The 16S sequences were filtered to remove sequences that (i) were <1 kb or >2 kb in length or (ii) contained non-A, T, C, G characters at >1% of nucleotide positions; of the initial set of 64,376 genomes, there were 53,796 genomes with at least one 16S sequence that passed the filtering steps.

We used two approaches for classifying the 16S sequences. First, we used the SINA (SILVA Incremental Aligner) Search & Classify tool[34]. SINA classification is based on similarity to curated 16S sequences present the following databases: the SILVA database[35], the Ribosomal Database Project (RDP)[36], the Genome Taxonomy Database (GTDB)[37], the Living Tree Project (LTP)[38], and EMBL-EBI/ENA. Results from each database were "translated" into NCBI's preferred nomenclature (*i.e.*, the NCBI-designated "Scientific Name" for each taxon) using the taxon "thesaurus" from NCBI Taxonomy (names.dmp, https://ftp.ncbi.nlm.nih.gov/pub/taxonomy/). We then identified "consensus" species assignments by majority rule: classifications that were assigned to a sequence by at least 3/5 SINA databases were retained, and sequences for which there was no consensus taxon we considered to be unclassified. Because databases sometimes provided a consensus taxon at one level but not others (*e.g.*, 3/5 databases agreed on the family assignment but there was no consensus at the order, class, or phylum levels), missing taxon information was filled in using the taxon level relationships (*i.e.*, connections between nodes on the bacterial phylogenetic tree)

currently accepted by NCBI Taxonomy (nodes.dmp, https://ftp.ncbi.nlm.nih.gov/pub/taxonomy/). Final SILVA/NCBI taxon assignments were inspected manually for accuracy of genus, family, order, class, and phylum relationships.

Because there was no consensus SINA-based taxon (consistent across 3/5 databases) for ~9% of the genomes (**Fig. S4**), we moved to a second method for 16S-based species classification. We assigned genomes to species by binning the genomes' 16S sequences into tentative species groups using the canonical 97% sequence identity cutoff using CD-Hit (cd-hit, https://github.com/weizhongli/cdhit/wiki)[39]. 16S clusters were designated with the genus most commonly assigned to the associated genomes by GTDB-Tk. Each genome contained an average of three 16S genes, and 16S-based assignments were not always consistent. In 1.2% of cases, 16S sequences from the same genome did not share at least 97% identity, and in ~9% of genomes, 16S-based genus assignments did not match whole-genome annotations (**Table S5**). We addressed this by removing 16S sequences for which the associated genome had a different genus assignment than the associated 16S cluster. Finally, a single genome was selected from each species cluster as the representative based on having (i) the 16S sequence with the fewest non-A, T, C, G characters, and in the case of a tie, (ii) the 16S sequence that was closest to 1550 bp, the median 16S length. This yielded a set of 5,568 representative genomes.

16S rRNA tree generation

16S sequences of the 5,568 16S cluster representatives were aligned using the MUSCLE v. 5.1 sequence alignment program (https://www.drive5.com/muscle/muscle.html)[40]. Five hundred diversified alignments were generated using the *super5* algorithm, each with a unique seed, and the best scoring alignment was identified by maximum column confidence score using the *maxcc* algorithm. From this highest scoring alignment, 2,000 uniquely seeded maximum likelihood trees were generated using RAxML-Next Generation (raxml-ng, https://github.com/amkozlov/raxml-ng)[41] using the GTR+G substitution model. To decrease processing time, columns in which >99% of sequences contained gaps were removed from the best-scoring alignment prior to tree generation. A final consensus tree was generated from the 2000 diversified trees using the *consense* algorithm by the Extended Majority Rule (MRE) method, and was visualized and annotated using the TreeViewer software (https://treeviewer.org/)[42].

For readability, the 5,568 16S cluster representatives were further consolidated into higher level groupings such that each group contained at least 5 genomes (which the exception of phyla for which fewer than 5 sequenced genomes are available, which remained single groups), resulting in 338 groups. Representatives for the 338 sequence groups were selected based on the number of genomes with >97% 16S sequence identity (*i.e.*, by largest cluster size), selecting only from sequences that were properly aligned on the initial 5,568 sequence tree. The representative sequences were aligned, with the addition of the 16S rRNA sequence from the archaeon *Thermococcus kodakarensis* KOD1 (NCBI Accession NR_028216.1) as the outgroup, and a consensus tree was generated as above. The final tree with 339 branches (338 bacterial groups and one 1 outgroup) was annotated to include the corresponding taxon for each branch and the number of genomes represented by each branch in TreeViewer.

Calculation and visualization of hopanoid biosynthesis gene conservation

We calculated the percentage of genomes within each 97% 16S identity cluster that were annotated in IMG to contain each hopanoid biosynthesis-associated protein domain or pathway from MetaCyc, TIGRFAM, Kegg Ontology, and Pfam (see **Table 1** for relevant IDs). The percentages for each domain or pathway were then averaged across clusters, and these average conservation values, along with taxon assignments, were mapped onto the 338-branch 16S rRNA sequence tree using TreeViewer.

Protein tree generation

Within our JGI/IMG genome dataset, we extracted all protein sequences containing the domains for SHC (KO:K06045), HpnH (IPR017833), HpnN (IPR017841), HpnP (KO:K22705) and HpnR (IPR027564). For computational efficiency, we used CD-HIT to remove highly similar sequences (>70% amino acid identity). For each protein family, sequences were aligned against outgroup sequences using MUSCLE v. 5.1. Outgroup sequences were selected as members of bacterial protein families that had manually reviewed annotations in the Swiss-Prot database (as indicated in UniprotKB). Because of their high functional and domain similarity, HpnR and HpnP sequences were included in the same alignment and serve as outgroups for one another. Additional outgroup information for each protein family follows: SHC – protein prenyltransferases (*e.g.*, protein farnesyltransferase subunit beta – Uniprot O13782) and bacterial di- and mono-terpene cyclases (*e.g.*, drimenyl disphosphate synthase (A0A2P2GK84), (+)-T-muurolol synthase (A7NH01)); HpnH – radical S-adenomethionine (rSAM) domain-containing synthases (*e.g.*, MoaA GTP 3',8-cyclase (P30745), biotin synthase (A0QX70), PqqA peptide cyclase (B0RQ25)); HpnP and HpnR – radical S-adenomethionine (rSAM) domain-containing methylases (*e.g.*, BchE (bacterial chlorophyll biosynthesis protein E), L-lysine 2,3-aminomutase (P39280)); HpnN – bacterial AcrB domain transporters (*e.g.*, Multidrug efflux pump subunit AcrB (P31224), Multidrug resistance protein MexB (P52002)).

For each protein family, 500 diversified alignments were generated using the *align* algorithm, each with a unique seed, and the best scoring alignment was identified by maximum column confidence score using *maxcc*. To decrease downstream processing time, columns in which >90% of sequences contained gaps were removed from the best-scoring alignment prior to tree generation. Maximum likelihood trees were generated using IQ-TREE 2 (https://iqtree.github.io/)[43]. First, the optimal nucleotide substitution model for each alignment was determined by ModelFinder[44] to be Q.pfam for all protein families. Using the optimal model for each, a maximum likelihood tree was generated using 1,000 replicates each for calculation of branch support values using ultrafast bootstrap approximation (UFBoot)[45] and SH-like approximate likelihood ratio test (sh-aLRT). Bayesian inference trees were generated using MrBayes 3.2.7a (https://nbisweden.github.io/MrBayes/)[46,47] with a mixed amino acid substitution model with gamma-distributed rate variation and a proportion of invariable sites. All other priors were set to default values and the algorithm was run for 10 million generations. As above, all trees were visualized and annotated using TreeViewer.

Calculation of protein domain correlations

Protein domain annotations for each bacterial genome were extracted through two approaches. For the full set of IMG genomes, gene annotations with Pfam, TIGRFAM, and KO IDs were downloaded directly from IMG. Because Pfam and TIGRFAM are legacy annotations that now have been integrated into the more current InterPro domain database, we also downloaded full genome sequences and annotated the genomes using the InterProScan command line tool (version 5.69-101.0, https://github.com/ebi-pf-team/interproscan)[48]. InterProScan analysis was only performed for the 5,568 cluster representatives to minimize processing needs.

We calculated correlation coefficients between pairs of domains by first determining the presence or absence of each Pfam, TIGRFAM, KO, and InterPro domain in the 5,568 cluster representative genomes. We then calculated the Pearson correlation coefficient and corresponding p-value for every pair of domains using binary presence (1) and absence (0) values. Correlation coefficients were calculated in custom Python scripts using the *scipy.stats.pearsonr* function.

B. diazoefficiens *nitrogen utilization assays*

Wild-type *Bradyrhizobium diazoefficiens* strain USDA110 spc4 and the cumate-inducible hopanoid mutant strain Pcu-shc::Δ*shc* (abbreviated as Δ*shc*) were cultured as described previously[23]. Briefly, an individual colony of each strain was incubated in liquid-rich AG liquid media aerobically in an incubating shaker at 30°C and 250 rpm until mid-log phase (OD_{600} = 0.4–0.8). WT, Δ*shc*, and Δ*shc* with cumate

supplementation (Δ*shc* + C) were master mixed with AG minimal media lacking a nitrogen source (2.3 mM sodium gluconate, 3.3 mM arabinose, 5.6 mM MES monohydrate, 5 mM HEPES, 1 mM Na2HPO4, 1.76 mM Na2SO4, 88 µM CaCl2, 25 µM FeCl3, and 0.73 mM MgSO4, pH 6.6) to an initial OD_{600} = 0.008. The master mix (100 µl) was dispensed into a 96-well Biolog PM3 Microplates containing 95 different nitrogen sources and a negative control. Each well was overlaid gently with 30 µL of mineral oil to prevent evaporation. The plate was placed in a BioTek microplate reader with continuous shaking, and the OD_{600} was measured every 2 h for 168 h. The growth curve data were plotted using Prism GraphPad.

Microsynteny analysis of shc- *and* hpnN-*adjacent genomic regions*

To compare regions surrounding *shc* loci across genomes, we first selected the subset of genomes containing *shc*. We defined *shc*-adjacent regions ("neighborhoods") as the nearest 20 upstream and 20 downstream neighboring ORFs. We used InterProScan to identify the protein domains encoded by each neighboring gene, and then we assigned each gene family a numerical ID based on the set of InterPro domains that it contained. For some genes, such as those known to be involved in hopanoid or isoprenoid biosynthesis, we replaced the numerical ID with a gene name such as "hpnN" or "hpnM". In some cases, genes contained some but not all domains of a known gene, and we assigned these genes the name of the known gene followed by an asterisk (*e.g.*, "hpnN*" or "hpnM*"). InterPro domains that occurred in fewer than 5% of the neighboring genes were excluded from this analysis, and genes containing only those domains were not assigned a name. To visualize microsynteny in these neighborhoods, we obtained the coordinates and orientation of each *shc*-neighboring gene and used a custom Python script to draw the genomic regions. To align the regions, the starting coordinate of *shc* was set to 0 for each genome. We colored each gene based on its assigned name/numerical ID and illustrated gene orientation using triangles pointed toward the 3' end.

To evaluate genes in the *hpnN* neighborhood, we selected genomes containing *hpnN* and removed any genomes that were not assigned to a Gram-negative species. We then calculated the percentage of genomes with *hpnM*, *hpnS*, and *hpnT* in the *hpnN* neighborhoods across genera. These values were visualized on a maximum likelihood tree of 16S rRNA sequences for species representatives of *hpnN*-containing, Gram-negative genera using TreeViewer.

HpnT, HpnM and HpnS structure and binding predictions

To predict the three-dimensional structures of HpnT, HpnM, and HpnS with diploptene, we used AlphaFold 3 (AF3, https://github.com/google-deepmind/alphafold3)[32]. For these predictions, we used the protein sequences of HpnM and HpnS from *Burkholderia multivorans* ATCC 17616 (BMULJ_05342 / Uniprot A0A0H3KU80 and BMULJ_05340 / Uniprot ID A0A0H3KPM6) and the protein sequence of HpnT from *Bradyrhizobium diazoefficiens* USDA110 (*bll3009* / Uniprot ID Q89QW5). The SMILES string of diploptene was obtained from PubChem. For HpnT, we predicted the structure of a HpnT monomer in complex with diploptene. For HpnM and HpnS, we predicted the structure of a HpnM / HpnS heterodimer in complex with diploptene. We used the default AF3 parameters for all predictions.

DISCUSSION

Our comparative genomics indicate that hopanoids most likely originated in an ancestor of the pesudomonadota, suggesting that the 2Me-geo-hopane index cannot be unambiguously used as markers of cyanobacterial expansion. This contradicts recent work by Hoshino and colleagues[14]. Although both studies find that *hpnP* is generally restricted to three phyla (pseudomonadota, cyanobacteriota, and methylomirabilota/rokubacteria), we differ in the degree of conservation of SHC and HpnP in cyanobacteria. We do not observe SHC or HpnP in the Synechococcales order, and we observe low levels of conservation of

HpnP within the aquatic heterocyst-forming Nostocaceae family. Our study uses more up-to-date taxonomic framework (GTDB) for our analysis, and differences in taxonomic classifications across genomes may account for some discrepancies. We also use stricter KO terms to identify HpnP and SHC homologs. Whereas Hoshino et al. report that SHC is widespread in bacillota based on identification of SHC-related proteins by BLAST, we posit that bacillota SHC homologs are mis-annotated SqhC proteins based on their phylogenetic relationship to experimentally verified SHC and SqhC proteins. Unfortunately, few HpnP homologs have been experimentally verified, and such validated reference points are needed for confident phylogenetics. We suggest that experimental studies to determine the substrate specificity of diverse HpnP orthologs will be required as benchmarks before true taxonomic range of *hpnP* can be determined.

Though we do not find that any hopanoid class can be linked to specific taxa, our analysis indicates that hopanoids are more strongly associated with high osmolarity environments and methanogenesis, methanotrophy, and nitrogen cycling. The link between hopanoids and osmotic stress tolerance is expected given the similarity of hopanoids to eukaryotic sterols, which are known to fortify cell membranes to reduce permeability and resist force-mediated lysis. It is well established that hopanoids similarly condense and thicken *in vitro* lipid bilayers[1,2], and though the relevance of these studies to living membranes is not completely clear, sensitivity to osmotic stress has been reported in nearly every hopanoid-deficient mutant tested[23,49-52]. We speculate that survival of high salinity environments was a major driver of hopanoid evolution.

The enrichment of single carbon- and nitrogen-metabolism in hopanoid-producing organisms is more difficult to parse. Many of these processes rely on isoprenoid quinones or radical SAM enzymes, and it is possible that their correlation with hopanoids arises from the use of shared substrates rather than a functional relationship. Presence of hopanoids in marine nitrite-oxidizing bacteria (NOB) has been recently emphasized in hopanoid geobiology literature[10-12], and in our growth assays, SHC is required for efficient nitrite utilization in the terrestrial nitrogen-fixing and denitrifying[53] alphaproteobacterium *Bradyrhizobium diazoefficiens*. Whether hopanoids participate directly in nitrite utilization will require future experiments. Though previous work identified an enrichment of hopanoids in plant-associated nitrogen-fixing bacteria[9], subsequent studies suggest that hopanoids do not play a direct role in nitrogen fixation[23,54]. Rather, we believe that hopanoids and nitrogen fixation are correlated because they are both common to low oxygen environments, given that hopanoids (unlike sterols) do not require oxygen for their biosynthesis and that oxygen is inhibitory to the nitrogen fixation machinery.

Beyond implications for geo-hopane interpretation, our work suggests new avenues for hopanoid cell biology. We identify four new uncharacterized but likely hopanoid-related proteins: the previously annotated HpnL and HpnM families and candidate proteins HpnS and HpnT, which are new to this study. Using the latest structure prediction tools, we predict that HpnS, HpnM, and HpnT participate in hopanoid trafficking across proteobacterial cell envelopes, likely interacting with the inner membrane hopanoid transporter HpnN in a manner analogous to the MlaC-MlaA-MlaBDEF phospholipid transport system. New analysis of *mlaC*-like gene copy numbers across bacteria[55] supports this notion. Tripathi et al. identified a subset of pseudomonadota that contain multiple copies of *mlaC*-like genes. In these species, which we find to be SHC-containing organisms, one *mlaC*-like copy is adjacent to other *mla* genes, whereas additional *mlaC*-like copies typically are adjacent to hopanoid-associated genes. The bacterial genera containing secondary, *hpnN*-adjacent *mlaC*-like genes in Tripathi et al. largely overlap with the genera in which we've identified *hpnM* genes, suggesting that we are referring to the same loci. This opens up future directions for cellular, structural, and biochemical assays to test this hypothetical transport pathway. If our functional predictions are verified, it will be interesting to study how these protein families originated, both relative to the Mla proteins and in the alpha- versus gamma-proteobacteria.

The function of HpnL remains unclear. Its closest relatives are the AglD family of lysylphosphatidylglycerol (lysyl-PG) synthases, which covalently link the carboxyl group of lysine to the

terminal hydroxyl in the glycerol moiety of PG[56,57], increasing the positive charge of the bilayer to improve resistance to cationic antimicrobial peptides. It is possible that HpnL might use a similar chemistry to join amino acids or other carboxylated compounds to the polyol groups of extended hopanoids such as bacteriohopanetetral (BHT), though the structure of BHT is quite distinct from the PG substrate of AglD. Covalent attachment of BHT to the amino acid derivative phenyl-acetic acid (PAA), an auxin-like compound and weak phytohormone, has been reported in actinobacterial Frankia spp.[58,59], suggesting other amino acid-hopanoid conjugates exist. However, the HpnL-associated protein domain (TIGR03476) is found predominantly in pseudomonadota, and to our knowledge PAA-BHT not been found in this phylum. The function of HpnL awaits future genetic characterization.

Finally, our work underscores the necessity of up-to-date taxonomic information for studies of gene distribution and evolution. By comparing the efficacy of the most used 16S rRNA-based taxon tools to the newer GTDB taxonomy, and calculating the frequency of mis-annotation due to 16S rRNA inconsistency (**Table S4**), we quantitatively demonstrate the importance of whole genome-based taxon assignments. We emphasize that caution should be used when relying on metagenomes for community assembly and gene evolution, as even the most unique and well-conserved markers such as 16S rRNA exhibit signs of horizontal transfer that are difficult to disentangle without whole genome data.

ACKNOWLEGEMENTS

This work was supported by NIH grants R00GM126141 and R35GM147015 (to B.J.B) and Carnegie Institution for Science endowment funds. Computational analyses were performed using Caltech's Resnick High Performance Computing Center and the NIH Biowolf HPC. We thank members of the Belin lab, Dr. Karina Gutiérrez-García (Carnegie/University of Arizona), Dr. Will Ludington (Carnegie/Johns Hopkins University), Dr. Damian Ekiert (Johns Hopkins University), and Dr. Dianne Newman (Caltech) for helpful discussions and feedback. We are grateful to Carnegie Embryology's IT, front office, and facilities support staff for making our work possible.

TABLES

Protein	Function	Pfam / Interpro	TIGRFam	Kegg Ontology	MetaCyc
SHC (HpnF)	C_{30} hopanoid production[60]	pfam13249/IPR032697, pfam13243/IPR032696, IPR006400	TIGR01507	KO:K06045	PWY-7072
HpnG	production of C_{35} hopanoid ribosylhopane[61,62]	IPR017831	TIGR03468	-	-
HpnH	production of C_{35} hopanoid adenosylhopane[61,63]	IPR017833	TIGR03470	-	-
HpnI	production of C_{35} hopanoid N-acetylglucosaminyl bacteriohopanetetrol (BHT)[64]	IPR017835	TIGR03472	-	-
HpnJ	production of C_{35} hopanoid BHT cyclitol ether[64]	IPR017834	TIGR03471	-	-
HpnK	production of C_{35} hopanoid glucosaminyl BHT[64]	IPR017836	TIGR03473	-	-
HpnL	Unknown	-	TIGR03476	-	-
HpnM	Unknown	IPR017842	TIGR03481	-	-
HpnN	Inner membrane hopanoid transport[65,66]	IPR017841	TIGR03480	-	-
HpnO/ HpnQ	production of C_{35} hopanoid aminobacteriohopanetetrol[61]	IPR005814**	"HpnO" underGe		
HpnP	C-2 methylation of hopanoids[67]	IPR034530*	-	KO:K22705	-
HpnR	C-3 methylation of hopanoids[68]	IPR027564	TIGR04367	KO:K22704	-

Table 1. Hopanoid-associated protein domain families and pathways in the Pfam/InterPro, TIGRFam, Kegg Ontology, and MetaCyc databases. *IPR034530 also contains hpnP-*like* domains and is not specific for HpnP. **IPR005814 contains other aminotransferases and is not specific for HpnO/HpnQ. In the TIGRFAM database, an entry under hopanoid lipid biosynthesis pathway (GenProp0761) named HpnO is annotated with "VacJ-like lipoprotein"/PF04333. These terms do not refer to the same protein; use of HpnO/HpnQ to refer to the enzyme responsible for aminobacteriohopanetetrol production is the more common designation. The HpnO protein associated with GenProp0761 in TIGRFAM is likely the protein we name HpnS in this manuscript.

Terpenoid/isoprenoid biosynthesis & co-factors		
Function	*r*	*Examples*
Hopanoid synthesis	0.595	squalene-hopene cyclase, hopanoid biosynthesis associated radical SAM protein HpnH, hopanoid C-2 methylase, hopanoid C-3 methylase
Squalene synthesis	0.412	presqualene diphosphate synthase, squalene synthases HpnC & HpnD, hydroxysqualene synthase
Radical SAM enzymes	0.225	methylthioribose-1-phosphate isomerase, S-adenosyl-L-homocysteine hydrolase, ergothioneine biosynthesis proteins EgtABC, coenzyme PQQ biosynthesis enzyme PqqE
Other terpenoid/ isoprenoid synthesis	0.217	farnesyl-diphosphate farnesyltransferase, geranyl diphosphate 2-C-methyltransferase, 2-methylisoborneol synthase, germacradienol/geosmin synthase, xanthine dehydrogenase, menaquinone-9 beta-reductase
B vitamin synthesis	0.195	precorrin-3B synthase, cobalamin biosynthesis protein CobT, thiamine biosynthesis protein ThiS, phosphoribosylformylglycinamidine (FGAM) synthase
Sterol metabolism/ transport	0.176	cholesterol oxidase, cholesterol uptake porter CUP1 of Mce4, phospholipid/cholesterol/gamma-HCH transport system proteins
Cellular homeostasis pathways		
Function	*r*	*Examples*
Trehalose/starch synthesis	0.248	malto-oligosyltrehalose synthase, trehalose synthase, starch synthase (maltosyl-transferring), maltose alpha-D-glucosyltransferase / alpha-amylase
Osmoresponsive signaling	0.203	osmosensitive K+ channel His kinase sensor domain, K+-transporting ATPase (KDP operon) subunits, KDP operon response regulator KdpE and sensor histidine kinase KdpD
Divalent cation transport	0.198	copper resistance proteins, periplasmic divalent cation tolerance protein, HoxN/HupN/NixA family nickel transporter, high affinity Mn2+ porin
Saccharide transport/ modification	0.194	beta-1,3-glucanase, beta-glucan exporter, alpha-glucan phosphorylases, non-reducing end alpha-L-arabinofuranosidase, isoamylase, OprB family carbohydrate-selective porin, xylobiose transporters
Taxon-specific metabolisms		
Function	*r*	*Examples*
Single C, N source utilization	0.213	ammonia/methane monooxygenase, carbon-monoxide dehydrogenase
Pentose phosphate pathway	0.205	6-phosphogluconate dehydrogenase, xylulose-5-phosphate/fructose-6-phosphate phosphoketolase, fructose-1,6-bisphosphatase II / sedoheptulose-1,7-bisphosphatase
Methanogenesis	0.196	formylmethanofuran dehydrogenase, methenyltetrahydromethanopterin cyclohydrolase, Cof coenzyme F420 synthases
Aerobic respiration	0.195	NADH:quinone oxidoreductases, cytochrome cbb3 oxidase, Cta/Caa/Cox-type cytochrome c oxidases
Glyoxylate pathway	0.195	glycolate oxidase, phosphoenolpyruvate guanylyltransferase, phosphoenolpyruvate phosphomutase
Nitrogen fixation	0.177	nif-specific ferredoxin, nitrogen fixation proteins NifN, NifT, NifW, NifX, FixT, FixU

Table 2. Protein domain categories positively correlated with SHC. Manually annotated domain categories correlated with SHC domain KO:K06045 across all genomes. Center column represents average correlation coefficient (r) for domains within each category. A full list of correlated domains is provided in Table S2.

FIGURE LEGENDS

Figure 1. Conservation of hopanoid biosynthesis-associated protein domains in sequenced bacteria. Consensus cladogram from 2,000 maximum likelihood trees generated from an alignment of representative bacterial 16S sequences from fully sequenced bacteria genomes in the JGI/IMG database. Each branch represents data from a group of phylogenetically related sequences; the number of genomes represented by each branch and the taxonomic classification of those genomes are provided as branch tip labels and branch names, respectively. Phyla assigned to genomes in each branch using the GTDB-Tk program are shown in the outer ring, with corresponding color-coding of the individual branches. Inner rings with grayscale rectangles indicate the degree of conservation within each branch of various hopanoid-associated domains and pathways from MetaCyc, Kegg Ontology, and TIGRFAM (see legend for domain/pathway represented by each ring). Scale bar indicates substitutions per 1000 nucleotide positions.

Figure 2. Hopanoid biosynthesis likely originated in the pseudomonadota. Maximum likelihood phylogenetic trees for protein sequences associated with SHC (KO:K06045), HpnH (IPR017833), HpnN (IPR017841), HpnP (KO:K22705) and HpnR (IPR027564). Branch names include likely taxa and JGI/IMG Genome ID. Numbers at internal nodes (highlighted in white) indicate branch support values using ultrafast bootstrap approximation (UFBoot), based on 1000 bootstraps; 100 = highest confidence, 0 = no confidence. Outgroup sequences are shown as grey triangles. Scale bar indicates substitutions per 1000 amino acids.

- *This figure has been compressed for bioRxiv compatibility. The full-size figure (~50 MB) is available here:* https://figshare.com/projects/hopanoid-distribution-supplemental-datasets/258284

Figure 3. The HpnM/S module and HpnT appear in distinct HpnN-encoding taxa. Maximum likelihood tree of 16S rRNA sequences from selected HpnN-encoding, Gram-negative bacterial genera. Branch tip labels (shaded rectangles) indicate conservation of HpnM, HpnS, and HpnT across all HpnN-encoding genomes within each genus; conservation levels corresponding to each shade are provided in the legend. Presence of these proteins was not annotated in the archaeon outgroup, *Thermococcus*. Scale bar indicates substitutions per 1000 nucleotides.

Figure 4. Candidate hopanoid transport pathways in proteobacteria. (**A**) Schematic of the Mla phospholipid (PL) transport system from *E. coli*. The inner membrane transport complex MlaBDEF transports PLs to the periplasmic shuttle protein MlaC, which interacts with OmpF/C-associated lipoprotein MlaA in the outer membrane. (**B**) Candidate hopanoid transport pathways in diderm bacteria. In most proteobacteria (left), the inner membrane transporter HpnN may interact with the periplasmic shuttle HpnM to deliver hopanoids to the outer membrane lipoprotein HpnS. In the Rhizobiales order (right), hopanoids trafficked by HpnN are delivered to the periplasmic HpnT protein. (**C**) AlphaFold 3-predicted structures of the *Burkholderia multivorans* HpnSM dimer (top) and *Bradyrhizobium diazoefficiens* HpnT (bottom) in association with the hopanoid diploptene (grey).

EXTENDED DATA LEGENDS

Extended Data Figure 1. Maximum likelihood tree of proteins annotated with both the N- and C-terminal domain PF IDs for SHC (pfam13249 and pfam13243). Colors at branch tips indicate GTDB-Tk assigned phyla for the genome where each protein is encoded; see Legend for color codes. Branches highlighted with arrows indicate seeded protein sequences for *B. subtilis* SqhC (Uniprot ID Q793C3), *Alicyclobacillus acidocaldarius* SHC (Uniprot ID P33247), and various eukaryotic lanosterol cyclases.

Extended Data Figure 2. Bayesian phylogenetic tree for protein sequences associated with SHC (KO:K06045). Branch names include likely taxa (generally at the genus level) and JGI/IMG Genome ID. Outgroup sequences are the same as in Fig. 2 shown as grey triangles. Scale bar indicates substitutions per 1000 amino acids.

Extended Data Figure 3. Modified bac120 bacterial species tree from GTDB (release 226, updated April 2025). Non-SHC-containing phyla were pruned, and only a single representative is shown for each bacterial family. Branches are labeled with GTDB family names only. Scale bar indicates substitutions per 1000 amino acids.

Extended Data Figure 4. Candidate hopanoid-associated genes *hpnL, hpnM, hpnS*, and *hpnT* are enriched near *shc* loci. Genes adjacent to the *shc* locus in selected bacterial genomes. Color-coding of ORFs indicates functional categories for conserved gene families: known hopanoid-associated genes (Hopanoid, yellow); known isoprenoid biosynthesis genes (Isoprenoid, pink), and candidate gene families *hpnM*, *hpnS*, *hpnT*, and *hpnL* (purple, green, blue, and orange, respectively). ORFs in the Other category (grey) are conserved gene families with unknown functions; number labels correspond to specific protein domain annotations provided in Table S3. Gene names labeled with asterisks indicate likely gene family members that lack one or more domain annotation relative to other genes in the same family (*e.g.*, '*hpnM**' = '*hpnM-like*'). All unlabeled ORFs (white) are non-conserved genes.

Extended Data Table 1. Percentage of genomes within each taxon (at the rank of Family and higher) annotated with each hopanoid-associated protein domain in Table 1. Percentages are averages of the percentages across species-level (16S rRNA) clusters, to minimize the effect of over-representation of model species.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Comparison of squalene cyclase domain annotations from JGI/IMG. Consensus cladogram from 2,000 maximum likelihood trees generated from an alignment of representative bacterial 16S sequences from fully sequenced bacteria genomes in the JGI/IMG database. Each branch represents data from a group of phylogenetically related sequences; the number of genomes represented by each branch and the taxonomic classification of those genomes are provided as branch tip labels and branch names, respectively. Phyla assigned to genomes in each branch using the GTDB-Tk program shown in the outer ring, with corresponding color-coding of the individual branches. Inner rings with grayscale rectangles indicate the degree of conservation within each branch of various hopanoid-associated domains and pathways from MetaCyc, Kegg Ontology, and TIGRFAM (see legend for domain/pathway represented by each ring).

Figure S2. Hopanoid-deficient *B. diazoefficiens* have slow growth on nitrite. Growth curves of WT (WT+E), hopanoid-deficient (delta *shc* + E), and rescue (delta *shc* + C) strains of *B. diazoefficiens* on minimal media supplemented with indicated nitrogen sources.

Figure S3. **Signal peptides are prevalent in HpnL, HpnM, HpnS, and HpnT proteins.** Fraction of proteins each family predicted to have a signal peptide for periplasmic secretion by Phobius[69].

Figure S4. Comparison of phylum-level annotations for bacterial 16S sequences using the GTDB-Tk and SILVA/NCBI approaches. Consensus cladogram from 2,000 maximum likelihood trees generated from an alignment of the 16S rRNA sequences from the 5,568 16S cluster representatives. Concentric rings are color-coded by phyla assigned to each branch by the SILVA/NCBI method (inner ring) or GTDB-Tk (outer ring). Grey rectangles represent branches for which phyla could not be determined unambiguously.

SUPPLEMENTAL TABLES

Supplemental tables are available at https://figshare.com/projects/hopanoid-distribution-supplemental-datasets/258284

Table S1. Accession details for all JGI/IMG bacterial genomes used for analysis in this study (first tab), including taxa assigned to each genome by GTDB-Tk and 16S rRNA clustering (second tab).

Table S2. Top 1000 protein domain families positively correlated with each domain term associated with hopanoid biosynthesis, across all genomes and within specific taxa. Each tab provides correlations for a separate domain.

Table S3. Conserved ORFs adjacent to *shc* and *hpnN* in bacterial genomes. Gene family list only includes families present in >5% of *shc*- or *hpnN*-containing genomes. Genes indicated with '-' are not conserved. Genome names are the assigned genus and IMG accession number.

Table S4. Taxon assignments for 16S rRNA sequences from genomes in Table S1 based on 97% identify sequence clustering versus whole genome taxon assignments from GTDB-Tk.

REFERENCES

1 Belin, B. J. *et al.* Hopanoid lipids: from membranes to plant-bacteria interactions. *Nat Rev Microbiol* **16**, 304–315 (2018). https://doi.org/10.1038/nrmicro.2017.173

2 Mitchison-Field, L. M. & Belin, B. J. Bacterial lipid biophysics and membrane organization. *Curr Opin Microbiol* **74**, 102315 (2023). https://doi.org/10.1016/j.mib.2023.102315

3 Summons, R. E., Jahnke, L. L., Hope, J. M. & Logan, G. A. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* **400**, 554–557 (1999). https://doi.org/10.1038/23005

4 Newman, D. K., Neubauer, C., Ricci, J. N., Wu, C.-H. & Pearson, A. Cellular and Molecular Biological Approaches to Interpreting Ancient Biomarkers. *Annual Review of Earth and Planetary Sciences* **44**, 493–522 (2016). https://doi.org/https://doi.org/10.1146/annurev-earth-050212-123958

5 Kuypers, M. M. M., van Breugel, Y., Schouten, S., Erba, E. & Damsté, J. S. S. N2-fixing cyanobacteria supplied nutrient N for Cretaceous oceanic anoxic events. *Geology* **32**, 853–856 (2004). https://doi.org/10.1130/g20458.1

6 Pearson, A. & Rusch, D. B. Distribution of microbial terpenoid lipid cyclases in the global ocean metagenome. *The ISME Journal* **3**, 352–363 (2009). https://doi.org/10.1038/ismej.2008.116

7 Naafs, B. D. A., Bianchini, G., Monteiro, F. M. & Sánchez-Baracaldo, P. The occurrence of 2-methylhopanoids in modern bacteria and the geological record. *Geobiology* **20**, 41–59 (2022). https://doi.org/https://doi.org/10.1111/gbi.12465

8 Ricci, J. N., Michel, A. J. & Newman, D. K. Phylogenetic analysis of HpnP reveals the origin of 2-methylhopanoid production in Alphaproteobacteria. *Geobiology* **13**, 267–277 (2015). https://doi.org/10.1111/gbi.12129

9 Ricci, J. N. *et al.* Diverse capacity for 2-methylhopanoid production correlates with a specific ecological niche. *Isme j* **8**, 675–684 (2014). https://doi.org/10.1038/ismej.2013.191

10 Mueller, A. J., Daebeler, A., Herbold, C. W., Kirkegaard, R. H. & Daims, H. Cultivation and genomic characterization of novel and ubiquitous marine nitrite-oxidizing bacteria from the Nitrospirales. *The ISME Journal* **17**, 2123–2133 (2023). https://doi.org/10.1038/s41396-023-01518-6

11 Elling, F. J. *et al.* Vitamin B_{12}-dependent biosynthesis ties amplified 2-methylhopanoid production during oceanic anoxic events to nitrification. *Proceedings of the National Academy of Sciences* **117**, 32996–33004 (2020). https://doi.org/doi:10.1073/pnas.2012357117

12 Elling, F. J. *et al.* Marine and terrestrial nitrifying bacteria are sources of diverse bacteriohopanepolyols. *Geobiology* **20**, 399–420 (2022). https://doi.org/10.1111/gbi.12484

13 Cornejo-Castillo, F. M. & Zehr, J. P. Hopanoid lipids may facilitate aerobic nitrogen fixation in the ocean. *Proceedings of the National Academy of Sciences* **116**, 18269–18271 (2019). https://doi.org/doi:10.1073/pnas.1908165116

14 Hoshino, Y. *et al.* Genetics re-establish the utility of 2-methylhopanes as cyanobacterial biomarkers before 750 million years ago. *Nature Ecology & Evolution* **7**, 2045–2054 (2023). https://doi.org/10.1038/s41559-023-02223-5

15 Wang, S., Meade, A., Lam, H.-M. & Luo, H. Evolutionary Timeline and Genomic Plasticity Underlying the Lifestyle Diversity in Rhizobiales. *mSystems* **5**, 10.1128/msystems.00438–00420 (2020). https://doi.org/10.1128/msystems.00438-20

16 Parks, D. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology* **36**, 996–+ (2018). https://doi.org/10.1038/nbt.4229

17 Chen, I.-M. A. *et al.* The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Research* **49**, D751–D763 (2020). https://doi.org/10.1093/nar/gkaa939

18 Chaumeil, P., Mussig, A., Hugenholtz, P. & Parks, D. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022). https://doi.org/10.1093/bioinformatics/btac672
19 Poralla, K., Hartner, T. & Kannenberg, E. Effect of temperature and pH on the hopanoid content of Bacillus acidocaldarius. *FEMS Microbiology Letters* **23**, 253–256 (1984).
20 Ochs, D., Kaletta, C., Entian, K. D., Beck-Sickinger, A. & Poralla, K. Cloning, expression, and sequencing of squalene-hopene cyclase, a key enzyme in triterpenoid metabolism. *J Bacteriol* **174**, 298–302 (1992). https://doi.org/10.1128/jb.174.1.298-302.1992
21 Parks, D. *et al.* A complete domain-to-species taxonomy for Bacteria and Archaea. *Nature Biotechnology* **38**, 1079–+ (2020). https://doi.org/10.1038/s41587-020-0501-8
22 Tookmanian, E. *et al.* Hopanoids Confer Robustness to Physicochemical Variability in the Niche of the Plant Symbiont Bradyrhizobium diazoefficiens. *J Bacteriol* **204**, e0044221 (2022). https://doi.org/10.1128/jb.00442-21
23 Pan, H., Shim, A., Lubin, M. B. & Belin, B. J. Hopanoid lipids promote soybean-Bradyrhizobium symbiosis. *mBio* **15**, e0247823 (2024). https://doi.org/10.1128/mbio.02478-23
24 Barison, N. *et al.* Protein HP1028 from the human pathogen Helicobacter pylori belongs to the lipocalin family. *Acta Crystallogr D Biol Crystallogr* **69**, 1387–1394 (2013). https://doi.org/10.1107/s0907444913008160
25 Yang, H. *et al.* Lipocalin family proteins and their diverse roles in cardiovascular disease. *Pharmacology & Therapeutics* **244** (2023). https://doi.org/10.1016/j.pharmthera.2023.108385
26 Malinverni, J. C. & Silhavy, T. J. An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane. *Proc Natl Acad Sci U S A* **106**, 8009–8014 (2009). https://doi.org/10.1073/pnas.0903229106
27 Ekiert, D. C. *et al.* Architectures of Lipid Transport Systems for the Bacterial Outer Membrane. *Cell* **169**, 273–285.e217 (2017). https://doi.org/10.1016/j.cell.2017.03.019
28 Al Jaberi, S. *et al.* Lipocalin-2: Structure, function, distribution and role in metabolic disorders. *Biomedicine & Pharmacotherapy* **142** (2021). https://doi.org/10.1016/j.biopha.2021.112002
29 Coudray, N. *et al.* Structure of bacterial phospholipid transporter MlaFEDB with substrate bound. *Elife* **9** (2020). https://doi.org/10.7554/eLife.62518
30 Abellón-Ruiz, J. *et al.* Structural basis for maintenance of bacterial outer membrane lipid asymmetry. *Nat Microbiol* **2**, 1616–1623 (2017). https://doi.org/10.1038/s41564-017-0046-x
31 Chong, Z. S., Woo, W. F. & Chng, S. S. Osmoporin OmpC forms a complex with MlaA to maintain outer membrane lipid asymmetry in Escherichia coli. *Mol Microbiol* **98**, 1133–1146 (2015). https://doi.org/10.1111/mmi.13202
32 Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024). https://doi.org/10.1038/s41586-024-07487-w
33 Yeow, J., Luo, M. & Chng, S. S. Molecular mechanism of phospholipid transport at the bacterial outer membrane interface. *Nat Commun* **14**, 8285 (2023). https://doi.org/10.1038/s41467-023-44144-8
34 Pruesse, E., Peplies, J. & Glöckner, F. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829 (2012). https://doi.org/10.1093/bioinformatics/bts252
35 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590–D596 (2012). https://doi.org/10.1093/nar/gks1219
36 Cole, J. R. *et al.* Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* **42**, D633–642 (2014). https://doi.org/10.1093/nar/gkt1244

37 Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research* **50**, D785–D794 (2021). https://doi.org/10.1093/nar/gkab776

38 Yilmaz, P. *et al.* The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* **42**, D643–648 (2014). https://doi.org/10.1093/nar/gkt1209

39 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012). https://doi.org/10.1093/bioinformatics/bts565

40 Edgar, R. C. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nature Communications* **13**, 6968 (2022). https://doi.org/10.1038/s41467-022-34630-w

41 Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019). https://doi.org/10.1093/bioinformatics/btz305

42 Bianchini, G. & Sánchez-Baracaldo, P. TreeViewer: Flexible, modular software to visualise and manipulate phylogenetic trees. *Ecol Evol* **14**, e10873 (2024). https://doi.org/10.1002/ece3.10873

43 Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020). https://doi.org/10.1093/molbev/msaa015

44 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587–589 (2017). https://doi.org/10.1038/nmeth.4285

45 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**, 518–522 (2017). https://doi.org/10.1093/molbev/msx281

46 Altekar, G., Dwarkadas, S., Huelsenbeck, J. P. & Ronquist, F. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20**, 407–415 (2004). https://doi.org/10.1093/bioinformatics/btg427

47 Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**, 539–542 (2012). https://doi.org/10.1093/sysbio/sys029

48 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014). https://doi.org/10.1093/bioinformatics/btu031

49 Tookmanian, E. *et al.* Hopanoids Confer Robustness to Physicochemical Variability in the Niche of the Plant Symbiont Bradyrhizobium diazoefficiens. *Journal of Bacteriology* **204**, e00442–00421 (2022). https://doi.org/10.1128/jb.00442-21

50 Welander, P. V. *et al.* Hopanoids play a role in membrane integrity and pH homeostasis in Rhodopseudomonas palustris TIE-1. *J Bacteriol* **191**, 6145–6156 (2009). https://doi.org/10.1128/jb.00460-09

51 Silipo, A. *et al.* Covalently linked hopanoid-lipid A improves outer-membrane resistance of a Bradyrhizobium symbiont of legumes. *Nature Communications* **5**, 5106 (2014). https://doi.org/10.1038/ncomms6106

52 Garby Tamsyn, J. *et al.* Lack of Methylated Hopanoids Renders the Cyanobacterium Nostoc punctiforme Sensitive to Osmotic and pH Stress. *Applied and Environmental Microbiology* **83**, e00777–00717 (2017). https://doi.org/10.1128/AEM.00777-17

53 Bedmar, E. J., Robles, E. F. & Delgado, M. J. The complete denitrification pathway of the symbiotic, nitrogen-fixing bacterium Bradyrhizobium japonicum. *Biochemical Society Transactions* **33**, 141–144 (2005). https://doi.org/10.1042/BST0330141

54 Belin, B. J., Tookmanian, E. M., de Anda, J., Wong, G. C. L. & Newman, D. K. Extended Hopanoid Loss Reduces Bacterial Motility and Surface Attachment and Leads to Heterogeneity in Root Nodule Growth Kinetics in a Bradyrhizobium-Aeschynomene Symbiosis. *Molecular Plant-Microbe Interactions* **32**, 1415–1428 (2019). https://doi.org/10.1094/MPMI-04-19-0111-R

55 Tripathi, R., Ayekpam, D. & Kanaujia, S. P. Unveiling multiple copies of MlaC highlights its multifaceted nature. *Archives of Microbiology* **207**, 107 (2025). https://doi.org/10.1007/s00203-025-04308-0

56 Staubitz, P., Neumann, H., Schneider, T., Wiedemann, I. & Peschel, A. MprF-mediated biosynthesis of lysylphosphatidylglycerol, an important determinant in staphylococcal defensin resistance. *FEMS Microbiol Lett* **231**, 67–71 (2004). https://doi.org/10.1016/s0378-1097(03)00921-2

57 Oku, Y., Kurokawa, K., Ichihashi, N. & Sekimizu, K. Characterization of the Staphylococcus aureus mprF gene, involved in lysinylation of phosphatidylglycerol. *Microbiology* **150**, 45–51 (2004). https://doi.org/https://doi.org/10.1099/mic.0.26706-0

58 Berry, A. M., Moreau, R. A. & Jones, A. D. Bacteriohopanetetrol: Abundant Lipid in Frankia Cells and in Nitrogen-Fixing Nodule Tissue 1. *Plant Physiology* **95**, 111–115 (1991). https://doi.org/10.1104/pp.95.1.111

59 Kleemann, G., Alskog, G., Berry, A. M. & Huss-Danell, K. Lipid composition and nitrogenase activity of symbioticFrankia (Alnus incana) in response to different oxygen concentrations. *Protoplasma* **183**, 107–115 (1994). https://doi.org/10.1007/BF01276818

60 Perzl, M., Müller, P., Poralla, K. & Kannenberg, E. L. Squalene-hopene cyclase from Bradyrhizobium japonicum: cloning, expression, sequence analysis and comparison to other triterpenoid cyclases. *Microbiology (Reading)* **143 (Pt 4)**, 1235–1242 (1997). https://doi.org/10.1099/00221287-143-4-1235

61 Welander, P. V. *et al.* Identification and characterization of Rhodopseudomonas palustris TIE-1 hopanoid biosynthesis mutants. *Geobiology* **10**, 163–177 (2012). https://doi.org/10.1111/j.1472-4669.2011.00314.x

62 Liu, W. *et al.* Ribosylhopane, a novel bacterial hopanoid, as precursor of C35 bacteriohopanepolyols in Streptomyces coelicolor A3(2). *Chembiochem* **15**, 2156–2161 (2014). https://doi.org/10.1002/cbic.201402261

63 Bradley, A. S., Pearson, A., Sáenz, J. P. & Marx, C. J. Adenosylhopane: The first intermediate in hopanoid side chain biosynthesis. *Organic Geochemistry* **41**, 1075–1081 (2010). https://doi.org/https://doi.org/10.1016/j.orggeochem.2010.07.003

64 Schmerk, C. L. *et al.* Elucidation of the Burkholderia cenocepacia hopanoid biosynthesis pathway uncovers functions for conserved proteins in hopanoid-producing bacteria. *Environ Microbiol* **17**, 735–750 (2015). https://doi.org/10.1111/1462-2920.12509

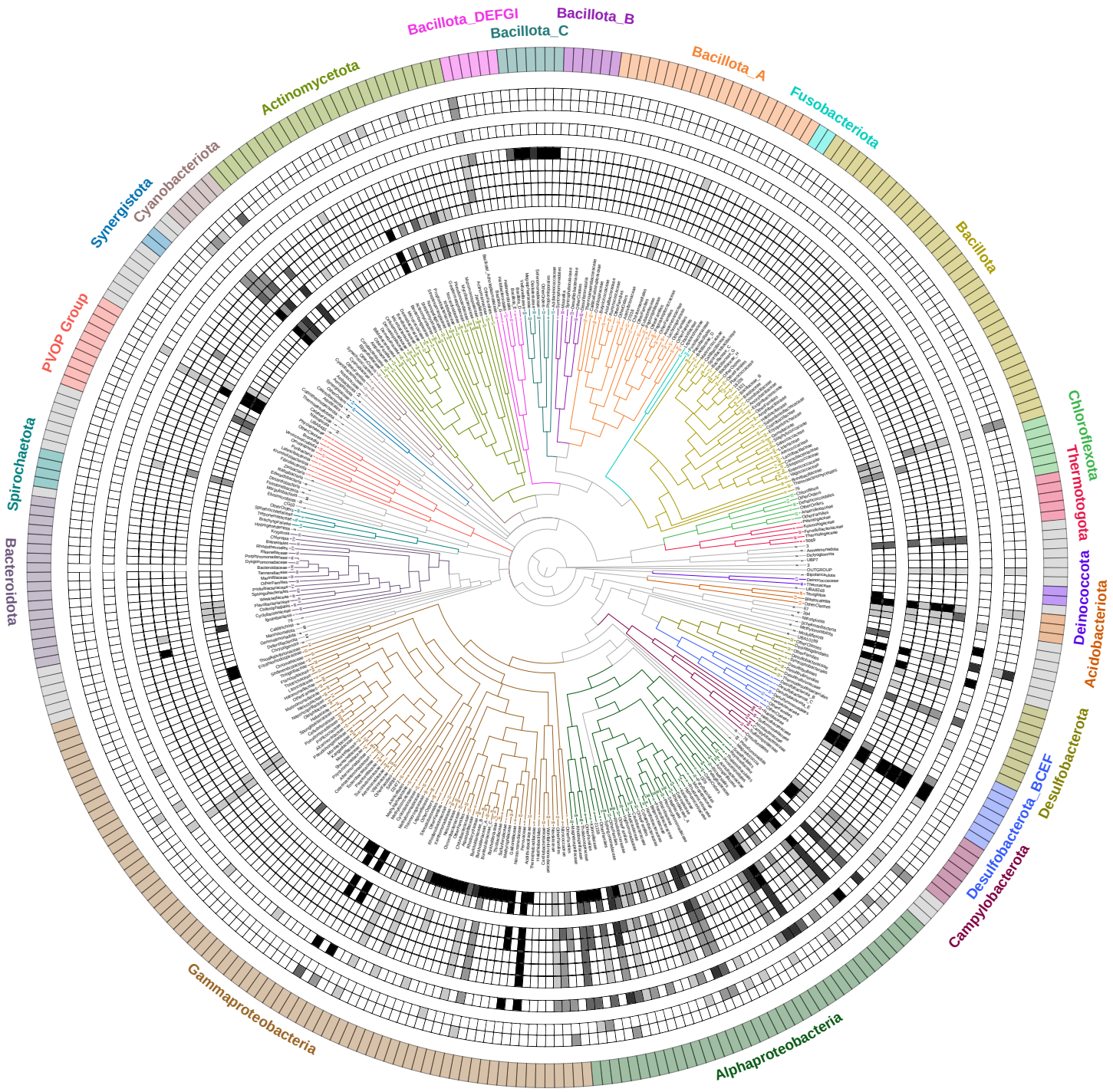
65 Doughty, D. M. *et al.* The RND-family transporter, HpnN, is required for hopanoid localization to the outer membrane of Rhodopseudomonas palustris TIE-1. *Proc Natl Acad Sci U S A* **108**, E1045–1051 (2011). https://doi.org/10.1073/pnas.1104209108

66 Kumar, N. *et al.* Crystal structures of the Burkholderia multivorans hopanoid transporter HpnN. *Proceedings of the National Academy of Sciences* **114**, 6557–6562 (2017). https://doi.org/doi:10.1073/pnas.1619660114

67 Welander, P. V., Coleman, M. L., Sessions, A. L., Summons, R. E. & Newman, D. K. Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes. *Proceedings of the National Academy of Sciences* **107**, 8537–8542 (2010). https://doi.org/doi:10.1073/pnas.0912949107

68 Welander, P. V. & Summons, R. E. Discovery, taxonomic distribution, and phenotypic characterization of a gene required for 3-methylhopanoid production. *Proc Natl Acad Sci U S A* **109**, 12905–12910 (2012). https://doi.org/10.1073/pnas.1208255109

69 Käll, L., Krogh, A. & Sonnhammer, E. L. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* **338**, 1027–1036 (2004). https://doi.org/10.1016/j.jmb.2004.03.016



Domain Conservation

- >80%
- >60%
- >40%
- >20%
- >0%
- Absent

Rings (Inner to Outer)

C30 Hopanoid Production

1. SHC (MetaCyc PWY-7072)
2. SHC (Kegg Ontology K06045)

C35 Hopanoid Modification

1. HpnH (TIGRFAM TIGR03470)
2. HpnG (TIGRFAM TIGR03468)
3. HpnI (TIGRFAM TIGR03472)
4. HpnJ (TIGRFAM TIGR03471)
5. HpnK (TIGRFAM TIGR03473)

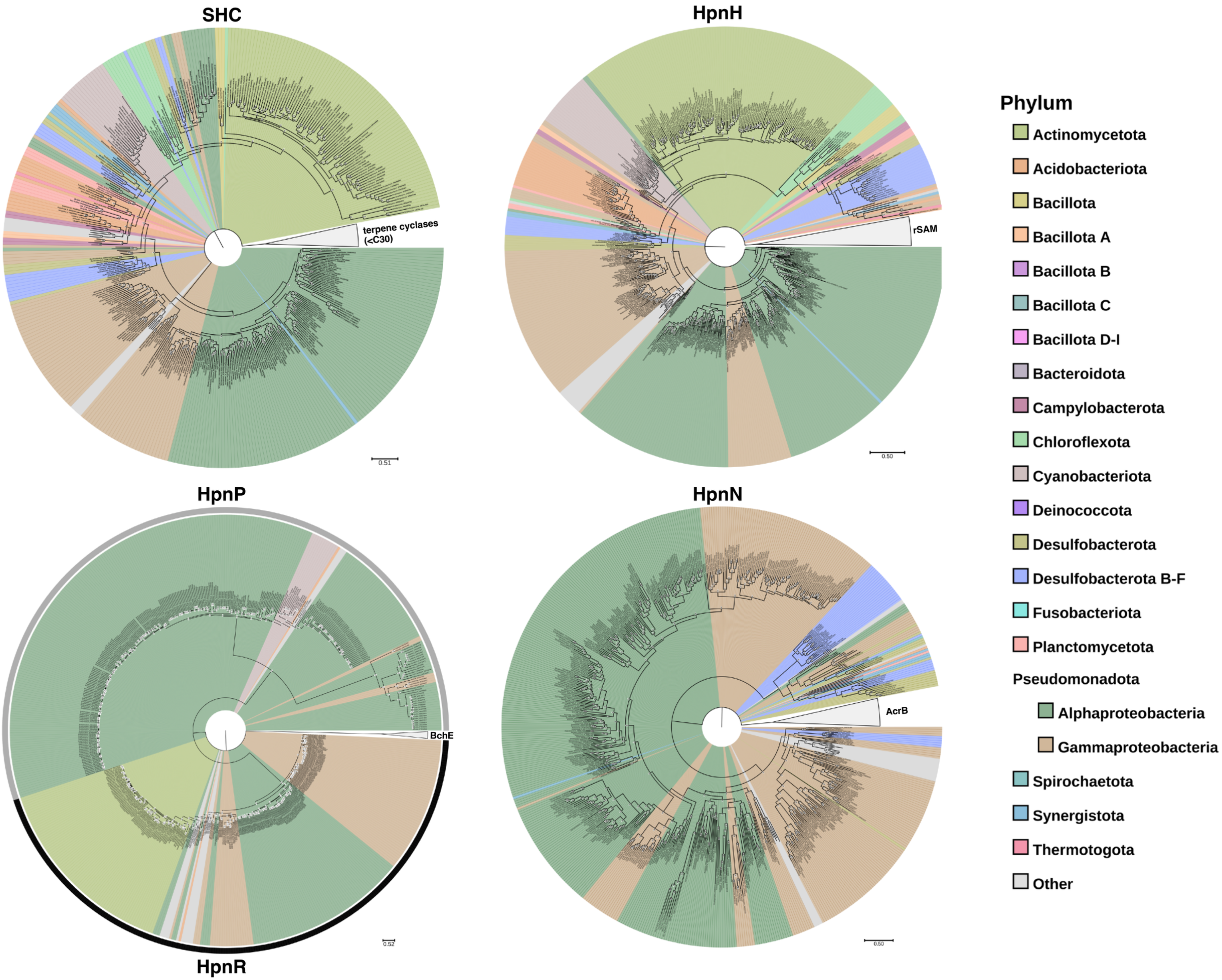
Hopanoid Transport

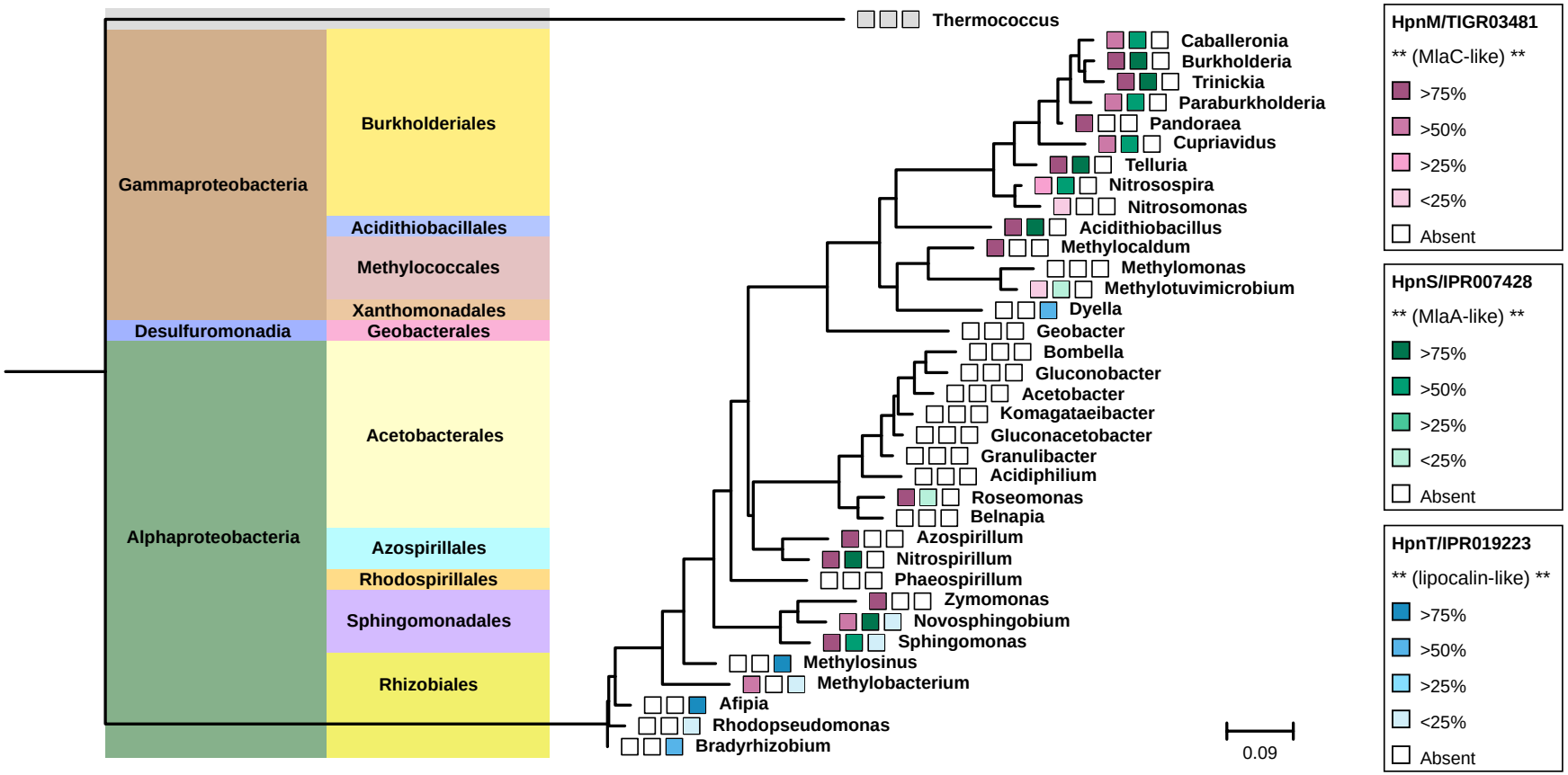
1. HpnN (TIGRFAM TIGR03480)

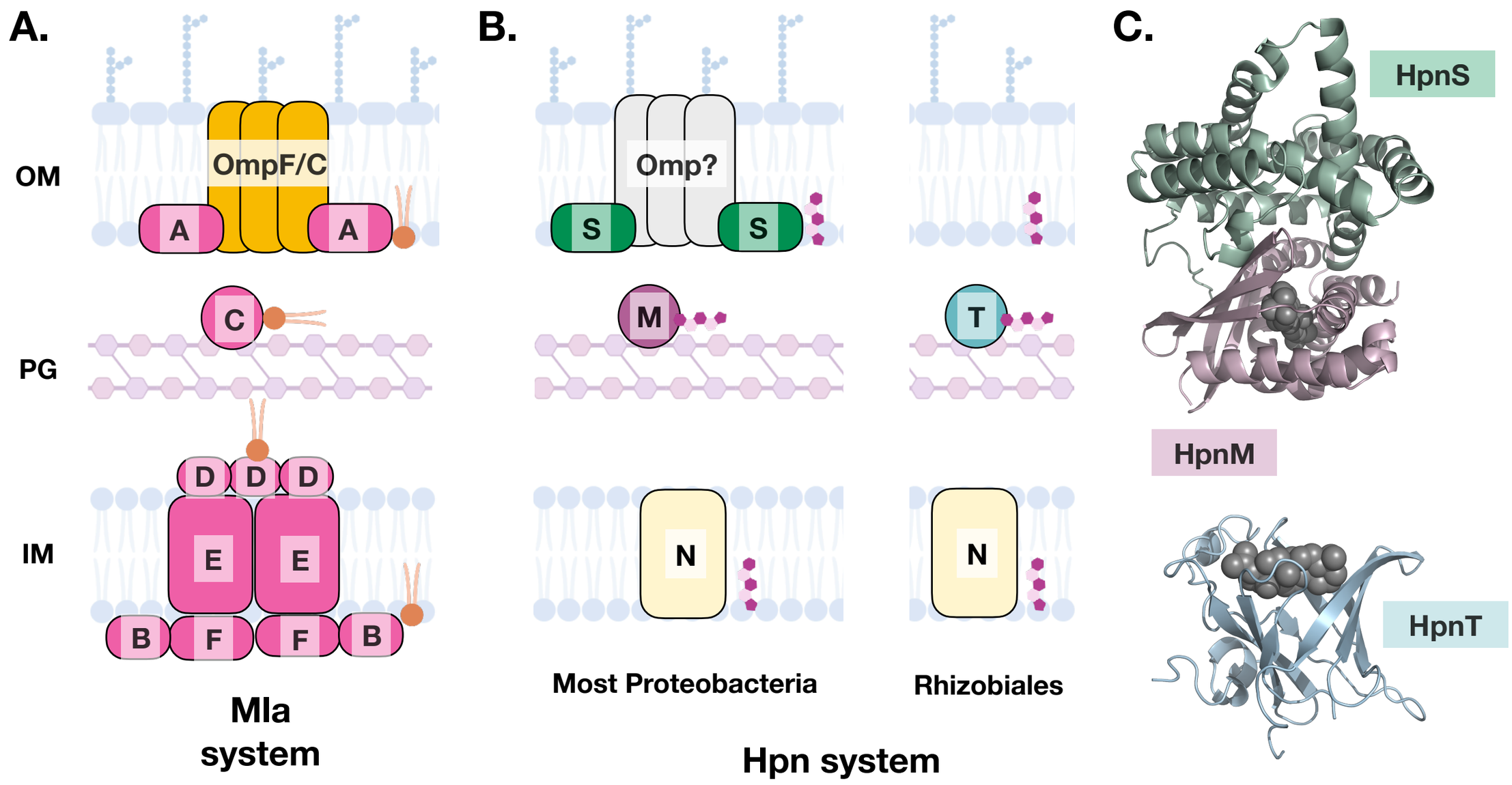
Hopanoid Methylation

1. HpnP (Kegg Ontology K22705)
2. HpnR (TIGRFAM TIGR04367)

13.0


SHC
terpene cyclases
(<C30)
0.51
HpnH
rSAM
0.50
HpnP
BchE
0.52
HpnN
AcrB
0.50
HpnR
Phylum
Actinomycetota
Acidobacteriota
Bacillota
Bacillota A
Bacillota B
Bacillota C
Bacillota D-I
Bacteroidota
Campylobacterota
Chloroflexota
Cyanobacteriota
Deinococcota
Desulfobacterota
Desulfobacterota B-F
Fusobacteriota
Planctomycetota
Pseudomonadota
Alphaproteobacteria
Gammaproteobacteria
Spirochaetota
Synergistota
Thermotogota
Other


Thermococcus
Caballeronia
Burkholderia
Trinickia
Paraburkholderia
Pandoraea
Cupriavidus
Telluria
Nitrosospira
Nitrosomonas
Acidithiobacillus
Methylocaldum
Methylomonas
Methylotuvimicrobium
Dyella
Geobacter
Bombella
Gluconobacter
Acetobacter
Komagataeibacter
Gluconacetobacter
Granulibacter
Acidiphilium
Roseomonas
Belnapia
Azospirillum
Nitrospirillum
Phaeospirillum
Zymomonas
Novosphingobium
Sphingomonas
Methylosinus
Methylobacterium
Afipia
Rhodopseudomonas
Bradyrhizobium
Gammaproteobacteria
Desulfuromonadia
Alphaproteobacteria
Burkholderiales
Acidithiobacillales
Methylococcales
Xanthomonadales
Geobacterales
Acetobacterales
Azospirillales
Rhodospirillales
Sphingomonadales
Rhizobiales
0.09
HpnM/TIGR03481
** (MlaC-like) **
>75%
>50%
>25%
<25%
Absent
HpnS/IPR007428
** (MlaA-like) **
>75%
>50%
>25%
<25%
Absent
HpnT/IPR019223
** (lipocalin-like) **
>75%
>50%
>25%
<25%
Absent

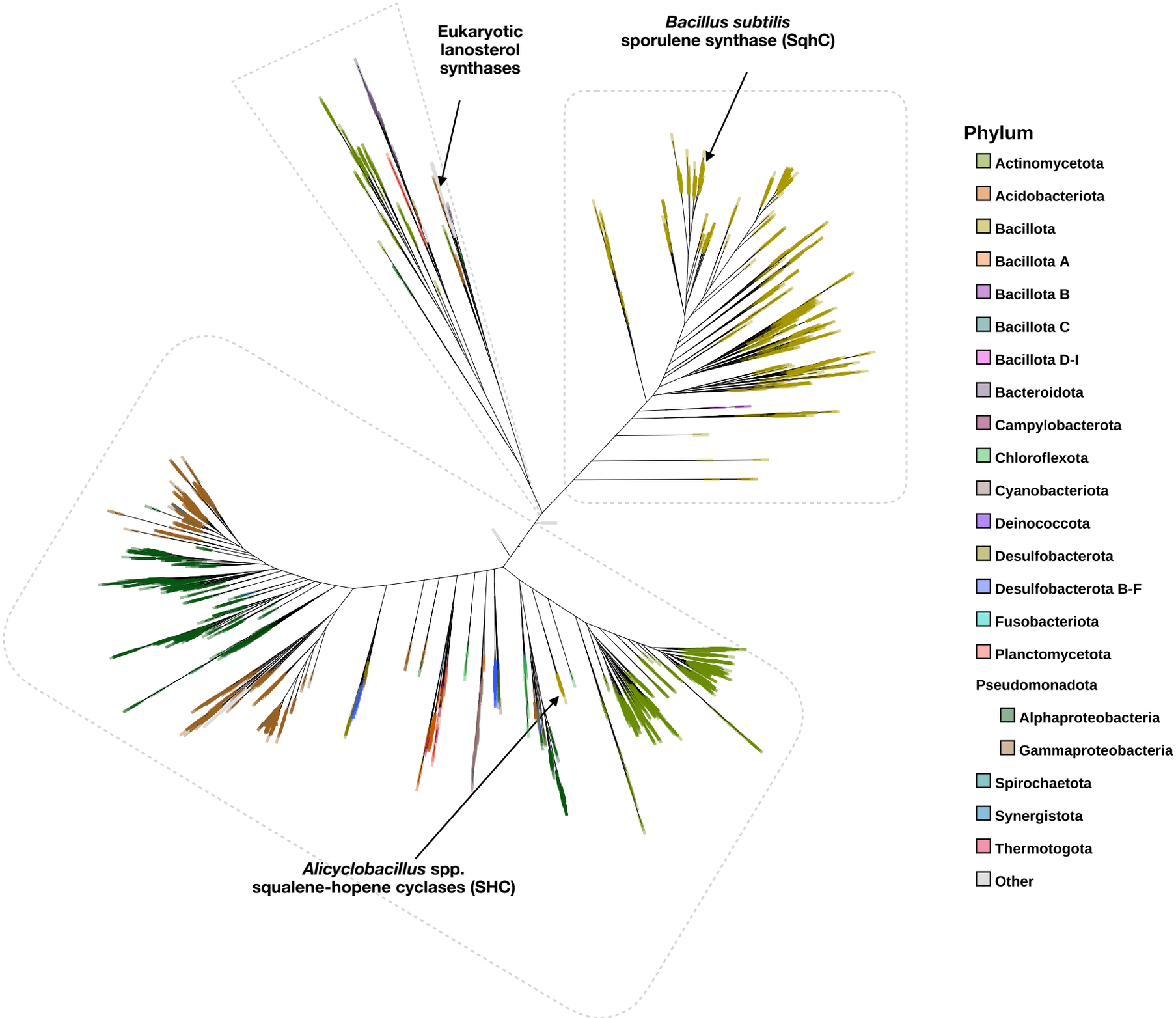

A.
OM
OmpF/C
A
A
C
PG
D
D
D
IM
E
E
B
F
F
B
Mla
system
B.
Omp?
S
S
M
T
N
N
Most Proteobacteria
Rhizobiales
Hpn system
C.
HpnS
HpnM
HpnT

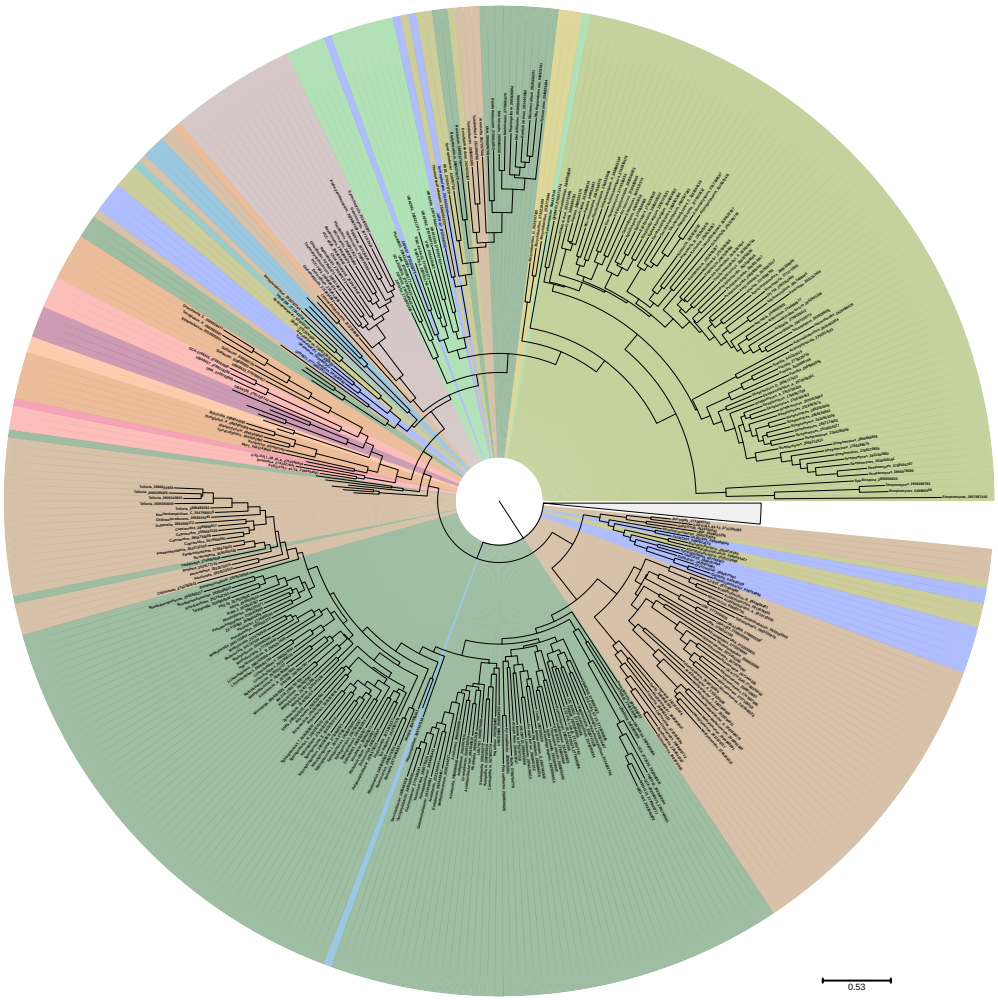
TAXON					C30 HOPANOIDS		C35 HOPANOIDS					TRANSPORT	METHYLATION	
Phylum	Class	Order	Family	Genomes	SHC_MC	SHC_KO	hpnH_TF	hpnG_TF	hpnJ_TF	hpnI_TF	hpnK_TF	hpnN_TF	hpnR_TF	hpnP_KO
Acidobacteriota														
		Blastocatellia		12	0.167	0.167	0.167	-	0.167	0.167	0.417	-	-	-
		Terriglobia		24	1.000	1.000	0.958	0.042	0.958	0.958	0.058	-	0.042	0.042
		Other		11	0.182	0.182	0.091	-	-	0.091	-	0.091	-	-
Actinomycetota														
	Acidimicrobiia													
		Acidimicrobiales		17	0.071	0.071	0.071	-	-	-	-	-	-	-
		Actinomarinales		33	-	-	-	-	-	-	-	-	-	-
	Actinomycetia													
		Actinomycetales												
			Dermatophilaceae	111	0.086	0.086	0.086	-	-	-	-	-	-	-
			Other	2085	-	-	-	-	-	-	-	-	-	-
		Jiangellales		19	-	-	-	-	-	-	-	-	-	-
		Mycobacteriales												
			Geodermatophilaceae	56	0.016	0.016	0.016	-	-	-	-	-	-	-
			Micromonosporaceae	180	0.206	0.206	0.175	-	-	-	-	-	-	-
			Mycobacteriaceae	4969	0.031	0.031	0.028	-	-	-	-	-	-	-
			Pseudonocardiaceae	250	0.539	0.539	0.459	-	-	-	-	-	0.019	-
		Nanopelagicales		33	-	-	-	-	-	-	-	-	-	-
		Propionibacteriales												
			Kribbellaceae	31	0.516	0.516	0.484	-	-	-	-	-	-	-
			Nocardioidaceae	168	0.006	0.006	0.006	-	-	-	-	-	-	-
			Propionibacteriaceae	260	-	-	-	-	-	-	-	-	-	-
		Streptomycetales		822	0.942	0.942	0.900	-	-	-	-	-	0.147	-
		Streptosporangiales		158	0.309	0.309	0.299	-	-	-	-	-	0.033	-
	Coriobacteriia			244	-	-	-	-	-	-	-	-	-	-
Aerophobota				4	-	-	-	-	-	-	-	-	-	-
Aquificota				19	-	-	-	-	-	-	-	-	-	-
Armatimonadota				2	0.500	0.500	0.500	-	0.500	0.500	0.500	-	-	-
Atribacterota				13	-	-	-	-	-	-	-	-	-	-
Bacillota														
	Bacilli													
		Acholeplasmatales		76	0.230	0.216	0.216	0.026	-	0.177	0.039	0.026	0.053	-
		Alicyclobacillales		76	0.230	0.216	0.216	0.026	-	0.026	0.039	0.026	0.053	-
		Bacillales												
			Anoxybacillaceae	101	-	-	-	-	-	-	-	-	-	-
			Bacillaceae	454	-	-	-	-	-	-	-	-	-	-
			Bacillaceae_B	18	-	-	-	-	-	-	-	-	-	-
			Bacillaceae_C	27	-	-	-	-	-	-	-	-	-	-
			Bacillaceae_G	759	-	-	-	-	-	-	-	-	-	-
			Bacillaceae_H	48	-	-	-	-	-	-	-	-	-	-
			Caldibacillaceae	10	-	-	-	-	-	-	-	-	-	-
			Domibacillaceae	17	-	-	-	-	-	-	-	-	-	-
			DSM-1321	37	-	-	-	-	-	-	-	-	-	-
			DSM-18226	45	-	-	-	-	-	-	-	-	-	-
			Planococcaceae	170	-	-	-	-	-	-	-	-	-	-
			Other	30	-	-	-	-	-	-	-	-	-	-
		Bacillales_D												
			Amphibacillaceae	99	-	-	-	-	-	-	-	-	-	-
			Halobacillaceae	24	-	-	-	-	-	-	-	-	-	-
			Other	18	-	-	-	-	-	-	-	-	-	-
		Bacillales_H												
			Bacillaceae_D	24	-	-	-	-	-	-	-	-	-	-
			Marinococcaceae	15	-	-	-	-	-	-	-	-	-	-
			Salisediminibacteriaceae	17	-	-	-	-	-	-	-	-	-	-
		Bacillales_K		18	-	-	-	-	-	-	-	-	-	-
		Brevibacillales		45	-	-	-	-	-	-	-	-	-	-
		Erysipelotrichales		108	-	-	-	-	-	-	-	-	-	-
		Exiguobacterales		58	-	-	-	-	-	-	-	-	-	-
		Lactobacillales												
			Carnobacteriaceae	74	-	-	-	-	-	-	-	-	-	-
			Other	5048	-	-	-	-	-	-	-	-	-	-
		Paenibacillales		285	-	-	-	-	-	-	-	-	-	-
		Staphylococcales		5823	-	-	-	-	-	-	-	-	-	-
		Thermoactinomycetales		28	-	-	-	-	-	-	-	-	-	-
		Other		39	-	-	-	-	-	-	-	-	-	-
Bacillota_A														
	Clostridia													
		Acetivibrionales												
			Acetivibrionaceae	16	-	-	-	-	-	-	-	-	-	-
			DSM-27016	13	0.200	0.200	-	-	-	-	-	-	-	-
		Christensenellales		11	-	-	-	-	-	-	-	-	-	-
		Clostridiales												
			Clostridiaceae	820	-	-	-	-	-	-	-	-	-	-
			Other	10	-	-	-	-	-	-	-	-	-	-
		Eubacteriales		23	-	-	-	-	-	-	-	-	-	-
		Lachnospirales												
			Lachnospiraceae	575	-	-	-	-	-	-	-	-	-	-
			Other	14	-	-	-	-	-	-	-	-	-	-
		Oscillospirales												
			Acutalibacteraceae	25	-	-	-	-	-	-	-	-	-	-
			Butyricicoccaceae	15	-	-	-	-	-	-	-	-	-	-
			Oscillospiraceae	48	-	-	-	-	-	-	-	-	-	-
			Ruminococcaceae	58	-	-	-	-	-	-	-	-	-	-
			Other	6	-	-	-	-	-	-	-	-	-	-
		Peptostreptococcales												
			Anaerovoracaceae	29	0.011	0.011	-	-	-	-	-	-	-	-

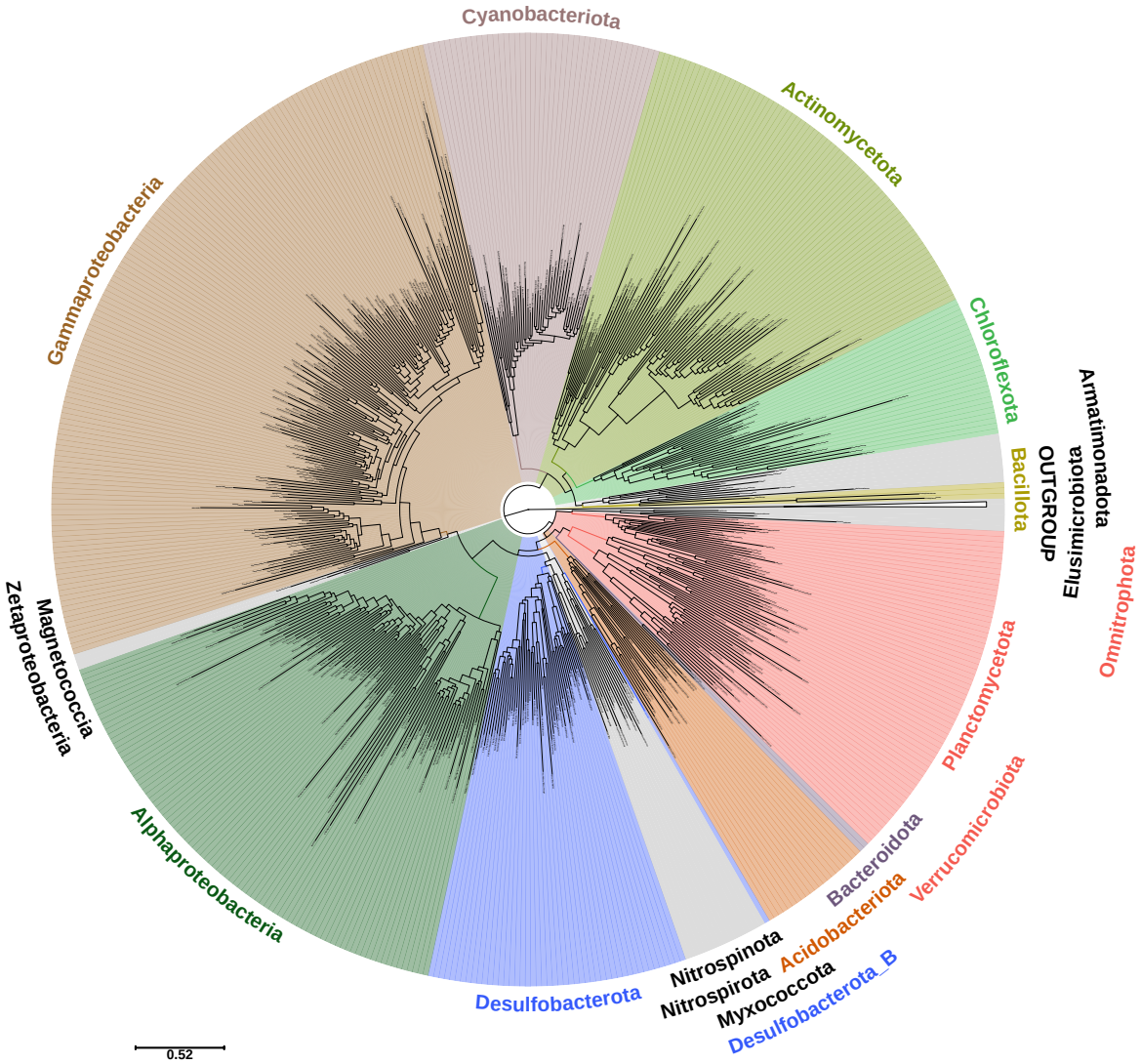
Peptostreptococcaceae	265	-	-	-	-	-	-	-	-	-	-
Other	29	-	-	-	-	-	-	-	-	-	-
Tissierellales											
Peptoniphilaceae	114	-	-	-	-	-	-	0.002	-	-	-
Other	27	-	-	-	-	-	-	-	-	-	-
Other	18	-	-	-	-	-	-	-	-	-	-
Thermoanaerobacteria	70	-	-	-	-	-	-	-	-	-	-
Bacillota_B											
Desulfitobacteriia											
Desulfitobacteriales											
Desulfitobacteriaceae	21	-	-	-	-	-	-	-	-	-	-
Syntrophobotulaceae	7	-	-	-	-	-	-	-	-	-	-
Desulfotomaculia	22	-	-	-	-	-	-	-	-	-	-
Moorellia	31	-	-	-	-	-	-	-	-	-	-
Syntrophomonadia	7	-	-	-	-	-	-	-	-	-	-
Other	15	-	-	-	-	-	-	-	-	-	-
Bacillota_C											
Negativicutes											
Acidaminococcales	17	-	-	-	-	-	-	0.833	-	-	-
Propionisporales	8	-	-	-	-	-	-	1.000	-	-	-
Selenomonadales	51	-	-	-	-	-	-	0.732	-	-	-
Veillonellales											
Dialisteraceae	10	-	-	-	-	-	-	0.929	-	-	-
Megasphaeraceae	32	-	-	-	-	-	-	1.000	-	-	-
Veillonellaceae	60	-	-	-	-	-	-	0.587	-	-	-
Other	12	-	-	-	-	-	-	0.833	-	-	-
Bacillota_DEF	46	-	-	-	-	-	-	-	-	-	-
Bacillota_G	5	-	-	-	-	-	-	0.400	-	-	-
Bacillota_I	76	0.230	0.216	0.216	0.026	-	0.026	0.039	0.026	0.053	-
Bacteroidota											
Bacteroidia											
Bacteroidales											
Bacteroidaceae	195	-	-	-	-	-	-	-	-	-	-
Dysgonomonadaceae	21	-	-	-	-	-	-	-	-	-	-
Marinifilaceae	26	-	-	-	-	-	-	-	-	-	-
Porphyromonadaceae	46	-	-	-	-	-	-	-	-	-	-
Prolixibacteraceae	24	-	-	-	-	-	-	-	-	-	-
Rikenellaceae	35	-	-	-	-	-	-	-	-	-	-
Tannerellaceae	69	-	-	-	-	-	-	-	-	-	-
Other	35	-	-	-	-	-	-	-	-	-	-
Chitinophagales	60	0.015	0.015	-	-	-	-	-	-	-	-
Cytophagales	59	-	-	-	-	-	-	-	-	-	-
Flavobacteriales											
Flavobacteriaceae	415	0.006	0.006	-	-	-	-	-	-	-	-
Weeksellaceae	49	0.012	0.012	-	-	-	-	-	-	-	-
Sphingobacteriales	118	-	-	-	-	-	-	-	-	-	-
Chlorobia	29	-	-	-	-	-	-	-	-	-	-
Ignavibacteria	7	0.714	-	-	-	-	0.143	-	-	-	-
Kryptonia	20	-	-	-	-	-	-	-	-	-	-
Rhodothermia	36	-	-	-	-	-	-	-	-	-	-
Bipolaricaulota	1	-	-	-	-	-	-	-	-	-	-
Caldisericota	1	-	-	-	-	-	-	-	-	-	-
Calditrichota	2	-	-	-	-	-	1.000	-	-	-	-
Calescibacterota	4	-	-	-	-	-	-	-	-	-	-
Campylobacterota											
Campylobacteria											
Campylobacterales	1167	-	-	-	-	-	-	-	-	-	-
Nautiliales	11	-	-	-	-	-	-	-	-	-	-
Desulfurellia	6	-	-	-	-	-	-	-	-	-	-
CG03	1	-	-	-	-	-	-	-	-	-	-
CG2-30-53-67	5	-	-	-	-	-	-	-	-	-	-
CG2-30-70-394	6	-	-	-	-	-	-	-	-	-	-
Chloroflexota											
Anaerolineae											
Anaerolineales											
Anaerolineaceae	13	-	-	-	-	-	-	-	-	-	-
Other	7	-	-	-	-	-	-	-	-	-	-
Other	9	0.111	-	-	-	-	-	-	-	-	-
Chloroflexia	12	0.083	0.083	-	-	-	-	-	-	-	-
Dehalococcoidia											
Dehalococcoidales	30	-	-	-	-	-	-	-	-	-	-
Other	15	0.267	0.267	0.267	-	-	-	-	-	-	-
Chrysiogenota	3	-	-	-	-	-	-	-	-	-	-
Coprothermobacterota	2	-	-	-	-	-	-	-	-	-	-
CSP1-3	2	-	-	-	-	-	-	-	-	-	-
Cyanobacteriota											
Cyanobacteriia											
Cyanobacteriales											
Cyanobacteriaceae	13	0.500	0.500	0.375	-	0.375	0.375	0.375	-	-	0.250
Nostocaceae	12	0.800	0.800	0.800	-	0.200	0.400	0.200	-	-	0.100
Other	29	0.724	0.724	0.759	-	0.207	0.517	-	-	-	0.034
PCC-6307	9	-	-	-	-	-	-	-	-	-	-
Synechococcales	5	-	-	-	-	-	-	-	-	-	-
Other	10	0.700	0.700	0.700	-	-	0.600	-	-	-	0.600
Deferribacterota	10	-	-	-	-	-	-	-	-	-	-
Deinococcota	63	-	-	-	-	-	-	-	-	-	-
Desantisbacteria	1	-	-	-	-	-	-	-	-	-	-
Desulfobacterota											
Desulfobacteria											

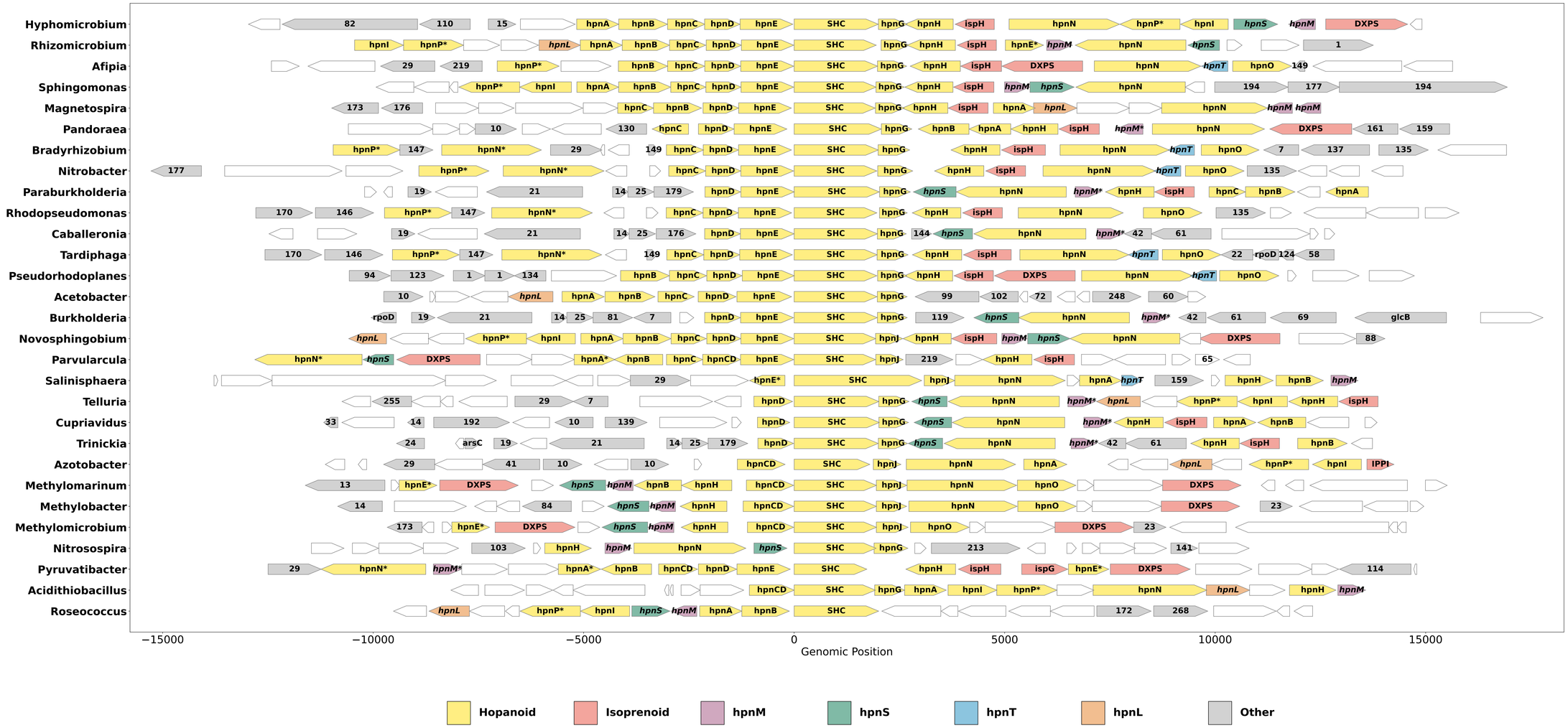
Desulfobacterales											
Desulfobacteraceae	9	-	-	-	-	-	-	-	-	-	-
Other	12	0.167	0.167	0.083	-	-	-	-	0.083	-	-
Desulfobulbia	19	-	-	-	-	-	-	0.058	-	-	-
Desulfovibrionia											
Desulfovibrionales											
Desulfomicrobiaceae	16	-	-	-	-	-	-	-	-	-	-
Desulfovibrionaceae	95	0.200	0.200	0.200	-	-	-	-	0.200	-	-
DSM-4 Desulfatiglandales	8	0.125	0.125	-	-	-	-	-	-	-	-
Syntrophobacteria	6	0.500	0.500	0.500	-	0.333	0.333	0.333	0.500	-	-
Thermodesulfobacteria	12	-	-	-	-	-	-	-	-	-	-
Other	12	0.083	0.083	0.083	-	0.083	0.083	0.083	-	-	-
Desulfobacterota_B	2	0.500	0.500	-	-	-	-	-	-	-	-
Desulfobacterota_C	1	-	-	-	-	-	-	-	-	-	-
Desulfobacterota_E	2	-	-	-	-	-	-	-	-	-	-
Desulfobacterota_F											
Desulfuromonadia											
Desulfuromonadales	18	0.222	0.222	0.222	-	-	-	0.083	0.222	-	-
Geobacterales											
Geobacteraceae	21	1.000	1.000	1.000	-	1.000	1.000	1.000	1.000	-	-
Other	9	1.000	1.000	1.000	-	1.000	1.000	1.000	1.000	-	-
Other	7	0.143	0.143	0.143	-	-	-	-	0.143	-	-
Dictyoglomota	2	-	-	-	-	-	-	-	-	-	-
Edwardsbacteria	6	-	-	-	-	-	-	-	-	-	-
Elusimicrobiota	11	0.273	0.273	0.273	-	0.091	-	-	-	-	-
Fibrobacterota	1	-	-	-	-	-	-	-	-	-	-
Firestonebacteria	2	-	-	-	-	-	-	-	-	-	-
Fusobacteriota	118	-	-	-	-	-	-	-	-	-	-
Gemmatimonadota	4	-	-	-	-	-	-	-	-	-	-
Hydrogenedentota	2	-	-	-	-	-	-	-	-	-	-
JdFR-76	1	-	-	-	-	-	-	-	-	-	-
Krumholzibacteriota	1	-	-	-	-	-	-	-	-	-	-
Latescibacterota	4	0.250	-	-	-	-	-	-	-	-	-
Margulisbacteria	10	-	-	-	-	-	-	-	-	-	-
Marinisomatota	10	-	-	-	-	-	-	-	-	-	-
Methylomirabilota	6	0.833	0.833	0.833	-	0.167	0.167	-	0.667	0.667	0.167
Moduliflexota	2	-	-	-	-	-	-	-	-	-	-
Myxococcota	76	0.230	0.216	0.216	0.026	-	0.026	0.039	0.026	0.053	-
Nitrospinota	2	0.500	0.500	-	-	-	-	-	1.000	-	-
Nitrospirota	11	0.091	-	-	-	-	0.227	-	-	-	-
Omnitrophota	27	0.074	0.074	0.074	-	-	-	-	-	-	-
Planctomycetota											
Brocadiia	5	1.000	1.000	1.000	-	0.150	0.150	0.150	-	-	-
Phycisphaerae	6	0.167	0.167	-	-	-	-	-	-	-	-
Other	5	1.000	1.000	0.200	-	-	-	-	0.200	-	-
Poribacteria	2	0.500	0.500	-	-	-	-	-	-	-	-
Pseudomonadota											
Alphaproteobacteria											
Acetobacterales	221	0.980	0.883	0.882	0.676	0.680	0.700	0.699	0.883	0.182	-
Azospirillales	61	1.000	0.290	0.290	0.189	0.091	0.091	0.091	0.290	-	-
Caulobacterales											
Caulobacteraceae	147	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.117	-	-
Hyphomonadaceae	25	-	-	-	-	-	-	-	-	-	-
Maricaulaceae	22	-	-	-	-	-	-	-	-	-	-
Other	11	0.312	0.312	0.312	0.208	-	-	-	0.312	-	-
Kiloniellales											
DSM-21159	5	0.667	0.667	0.667	0.667	0.333	0.333	0.333	0.667	0.333	-
Kiloniellaceae	5	-	-	-	-	-	-	-	-	-	-
Other	4	1.000	0.500	0.500	-	-	-	-	0.500	-	-
MED-G09	5	-	-	-	-	-	-	-	-	-	-
Micropepsales	5	1.000	1.000	1.000	1.000	0.750	0.750	0.750	0.750	-	-
Parvibaculales	11	0.273	0.182	0.273	-	-	-	-	0.409	-	-
Pelagibacterales	89	-	-	-	-	-	-	-	-	-	-
Puniceispirillales	15	-	-	-	-	-	-	-	-	-	-
Reyranellales	9	0.786	0.786	0.786	0.786	0.500	0.500	0.500	0.786	0.143	-
Rhizobiales											
Beijerinckiaceae	140	0.692	0.583	0.583	0.580	0.389	0.389	0.361	0.552	-	0.374
Cohaesibacteraceae	7	-	-	-	-	-	-	-	-	-	-
Devosiaceae	53	-	-	-	-	-	-	-	-	-	-
Hyphomicrobiaceae	14	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	-	0.083
Kaistiaceae	7	-	-	-	-	-	-	-	-	-	-
Phreatobacteraceae	5	-	-	-	-	-	-	-	-	-	-
Pleomorphomonadaceae	13	-	-	-	-	-	-	-	-	-	-
Rhizobiaceae	1650	0.014	0.014	0.012	0.006	-	-	-	0.012	-	0.016
Rhodomicrobiaceae	6	0.667	0.667	0.667	0.667	-	-	-	0.667	-	-
Stappiaceae	102	-	-	-	-	-	-	-	-	-	0.089
Xanthobacteraceae	234	0.647	0.633	0.633	0.598	0.233	0.234	0.234	0.633	-	0.400
Other	31	0.204	0.204	0.204	0.161	0.065	0.129	0.129	0.204	-	0.065
Rhodobacterales	950	0.053	-	-	-	-	-	-	-	-	-
Rhodospirillales											
Magnetospirillaceae	21	1.000	0.600	0.600	0.600	0.600	0.600	0.600	0.600	-	-
Rhodospirillaceae	19	1.000	0.600	0.600	0.600	-	-	-	0.600	-	0.020
Thalassospiraceae	26	1.000	-	-	-	-	-	-	-	-	0.333
Other	9	1.000	0.444	0.444	0.333	0.111	0.111	0.111	0.444	-	-
Rickettsiales	13	-	-	-	-	-	-	-	-	-	-
Sneathiellales	5	-	-	-	-	-	-	-	-	-	-
Sphingomonadales											
Sphingomonadaceae	738	0.121	0.121	0.113	0.036	0.112	0.112	0.104	0.113	-	-
Other	11	0.273	0.273	0.273	-	-	-	-	0.273	-	-

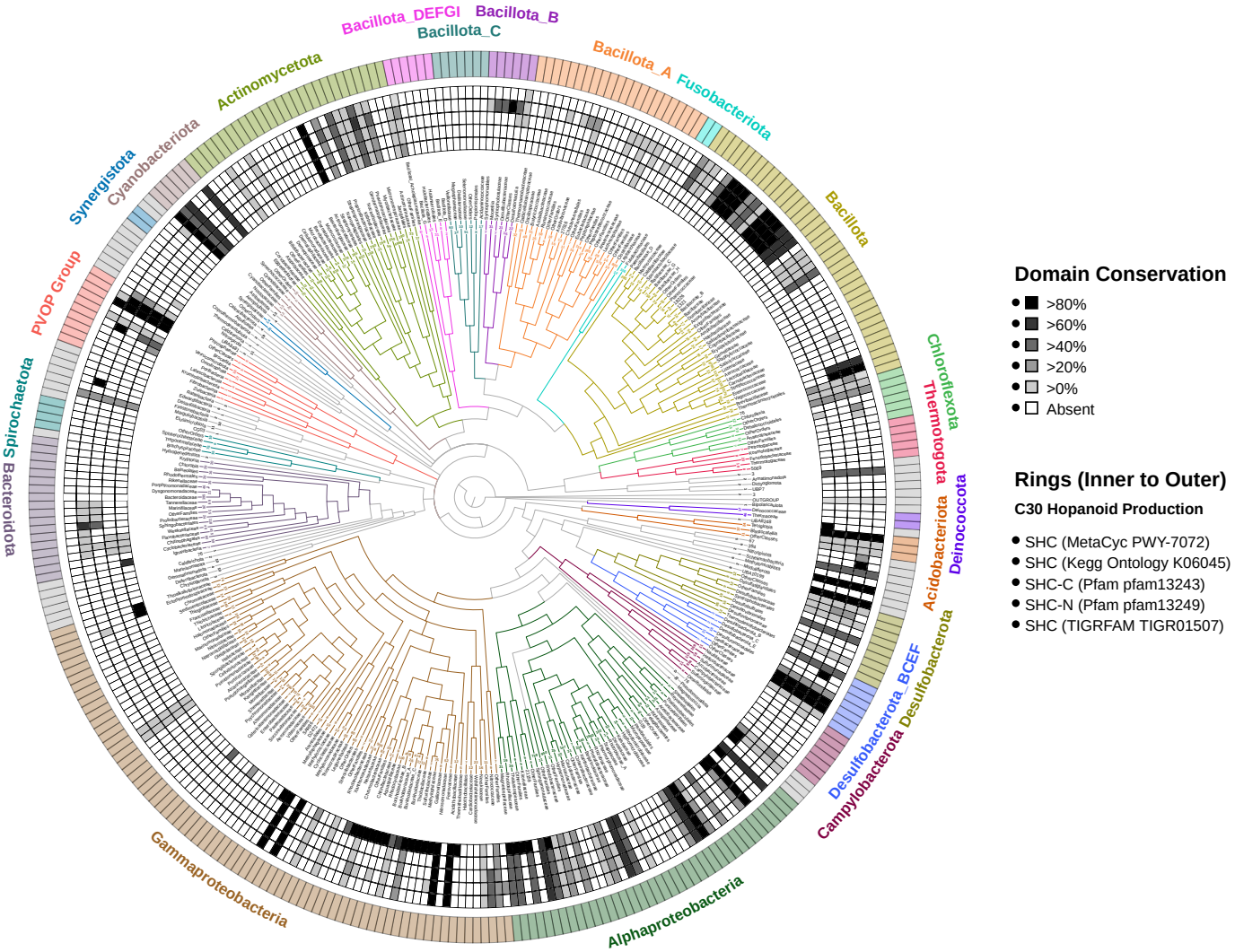
Thalassobaculales	6	1.000	-	-	-	-	-	-	-	-	0.250
Other	41	0.524	0.276	0.260	0.152	0.140	0.140	0.091	0.244	0.037	0.049
Gammaproteobacteria											
Acidithiobacillales											
Acidithiobacillaceae	12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-	-
Thermithiobacillaceae	5	-	-	-	-	-	-	-	-	-	-
Arenicellales	8	-	-	-	-	-	-	-	-	-	-
Burkholderiales											
Aquaspirillaceae	9	1.000	-	-	-	-	-	-	-	-	-
Burkholderiaceae	1713	0.838	0.357	0.365	0.345	0.286	0.295	0.307	0.364	0.096	-
Burkholderiaceae_A	26	0.417	0.083	0.083	0.083	0.083	0.083	0.083	0.083	-	-
Burkholderiaceae_B	458	0.872	-	-	-	-	-	-	-	-	-
Burkholderiaceae_C	474	0.646	-	-	-	-	-	-	-	-	-
Chitinibacteraceae	21	1.000	-	-	-	-	-	-	-	-	-
Chromobacteriaceae	83	0.993	-	-	-	-	-	-	-	-	-
Ferrovaceae	6	-	-	-	-	-	-	-	-	-	-
Gallionellaceae	8	1.000	-	-	-	-	-	-	-	-	-
Methylophilaceae	75	0.818	-	-	-	-	-	-	-	-	-
Neisseriaceae	550	0.562	-	-	-	-	-	-	-	-	-
Nitrosomonadaceae	52	1.000	1.000	1.000	1.000	-	-	-	1.000	-	-
Rhodocyclaceae	100	0.914	-	-	-	-	-	-	-	-	0.029
Sulfuriferulaceae	7	1.000	-	-	-	-	-	-	-	-	-
Thiobacillaceae	9	1.000	-	-	-	-	-	-	-	-	-
Other	20	0.900	0.050	0.050	0.050	-	-	-	0.050	-	-
Cardiobacteriales	34	-	-	-	-	-	-	-	-	-	-
Chromatiales											
Chromatiaceae	15	-	-	-	-	-	-	-	-	-	-
Sedimenticolaceae	14	-	-	-	-	-	-	-	-	-	-
Ectothiorhodospirales											
Ectothiorhodospiraceae	12	1.000	-	-	-	-	-	-	-	-	-
Thioalkalivibrionaceae	74	1.000	-	-	-	-	-	-	-	-	-
Enterobacterales											
Aeromonadaceae	254	0.332	-	-	-	-	-	-	-	-	-
Alteromonadaceae	390	-	-	-	-	-	-	-	-	-	-
Celerinatantimonadaceae	7	0.600	-	-	-	-	-	-	-	-	-
Shewanellaceae	123	-	-	-	-	-	-	-	0.063	-	-
Vibrionaceae	1501	-	-	-	-	-	-	-	0.026	-	-
Other	8168	-	-	-	-	-	-	-	-	-	-
Francisellales	146	-	-	-	-	-	-	-	-	-	-
Halothiobacillales	7	-	-	-	-	-	-	-	-	-	-
Legionellales	80	-	-	-	-	-	-	-	-	-	-
Methylococcales											
Cycloclasticaceae	12	-	-	-	-	-	-	-	-	-	-
Methylococcaceae	9	1.000	1.000	1.000	1.000	-	-	-	1.000	0.600	-
Methylomonadaceae	33	1.000	1.000	1.000	1.000	-	-	-	1.000	0.347	-
Nevskiales											
Nevskiaceae	16	0.083	0.083	0.083	0.042	0.083	0.083	0.083	0.250	-	-
Other	11	-	-	0.509	-	0.255	0.382	0.382	0.382	-	-
Nitrococcales											
Nitrococcaceae	9	0.250	0.250	0.250	0.250	-	-	-	0.250	0.250	-
Other	7	-	-	-	-	-	-	-	-	-	-
Nitrosococcales	12	1.000	-	-	-	-	-	-	-	-	-
PS1	25	-	-	-	-	-	-	-	-	-	-
Pseudomonadales											
Cellvibrionaceae	65	0.071	0.071	0.071	-	-	-	-	-	-	-
Pseudomonadaceae	2184	0.045	0.045	0.043	0.022	0.043	0.043	0.043	0.043	-	-
Other	2050	-	-	-	-	-	-	-	-	-	-
SAR86	17	-	-	-	-	-	-	-	-	-	-
Steroidobacterales	5	-	-	-	-	-	-	-	-	-	-
Thiomicrospirales	35	-	-	-	-	-	-	-	-	-	-
Thiotrichales	11	-	-	-	-	-	-	-	-	-	-
Xanthomonadales											
Rhodanobacteraceae	72	0.115	0.115	0.115	0.006	0.115	0.115	0.115	0.138	-	-
Other	725	-	-	-	-	-	-	-	-	-	-
Other	60	0.339	0.100	0.100	0.100	0.072	0.072	0.035	0.092	0.044	-
Magnetococcia	1	1.000	-	-	-	-	-	-	-	-	-
Zetaproteobacteria	33	0.929	0.076	0.133	0.019	0.057	0.057	0.057	0.076	-	-
Ratteibacteria	3	-	-	-	-	-	-	-	-	-	-
Schekmanbacteria	1	-	-	-	-	-	-	-	-	-	-
Spirochaetota	115	-	-	-	-	-	-	-	-	-	-
Synergistota	34	-	-	-	-	-	-	-	-	-	-
Thermodesulfobiota	1	-	-	-	-	-	-	-	-	-	-
Thermotogota	65	-	-	-	-	-	-	-	-	-	-
UBA10199	8	1.000	1.000	1.000	-	-	-	-	1.000	-	-
UBA8248	2	1.000	-	-	-	-	-	-	-	-	-
UBA8481	2	-	-	-	-	-	-	-	-	-	-
UBP7	1	-	-	-	-	-	-	-	-	-	-
Verrucomicrobiota	16	-	-	-	-	-	-	-	-	-	-
WOR-3	2	-	-	-	-	-	-	-	-	-	-
Zixibacteria	2	-	-	-	-	-	-	-	-	-	-

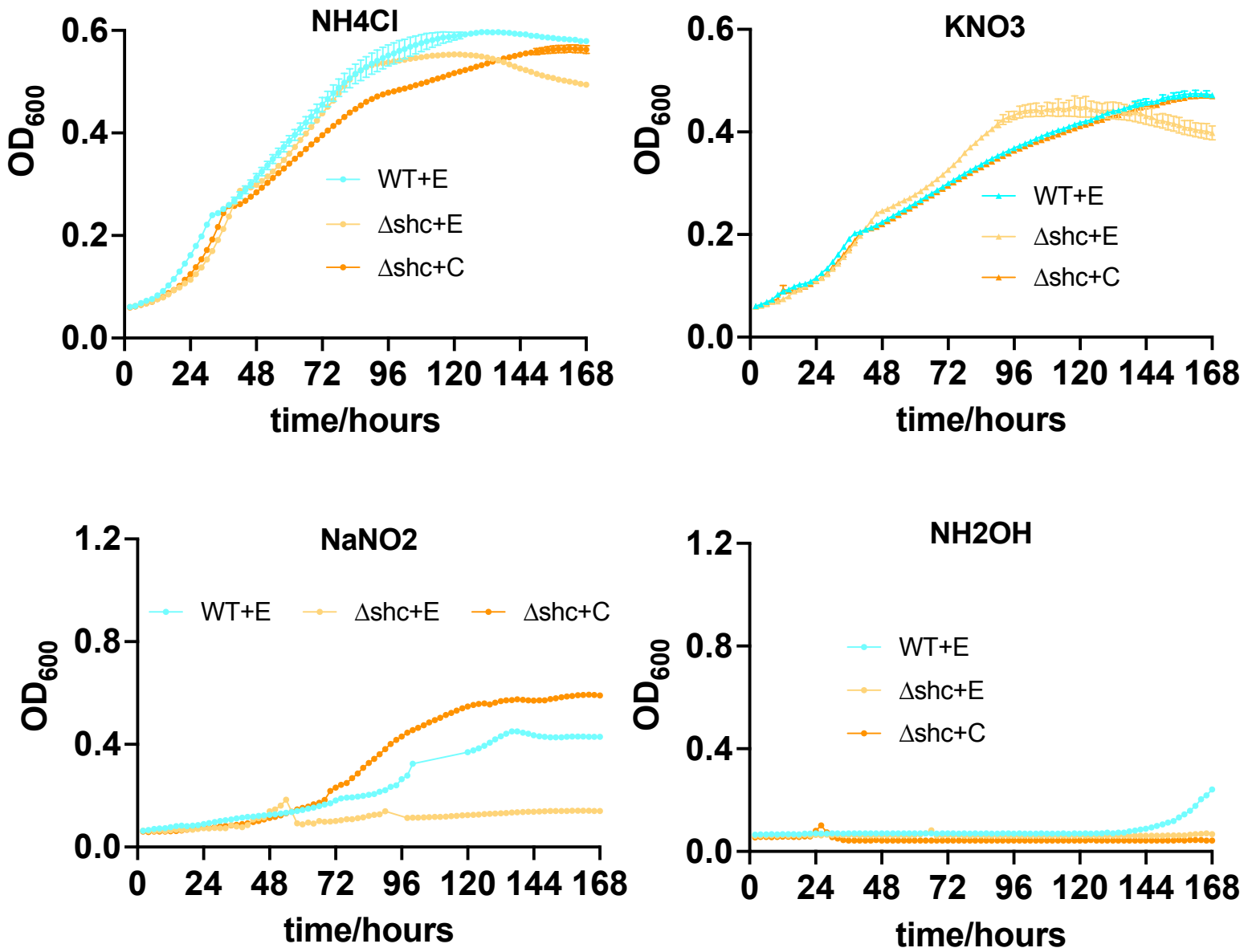

Eukaryotic lanosterol synthases
Bacillus subtilis sporulene synthase (SqhC)
Alicyclobacillus spp. squalene-hopene cyclases (SHC)
Phylum
Actinomycetota
Acidobacteriota
Bacillota
Bacillota A
Bacillota B
Bacillota C
Bacillota D-I
Bacteroidota
Campylobacterota
Chloroflexota
Cyanobacteriota
Deinococcota
Desulfobacterota
Desulfobacterota B-F
Fusobacteriota
Planctomycetota
Pseudomonadota
Alphaproteobacteria
Gammaproteobacteria
Spirochaetota
Synergistota
Thermotogota
Other

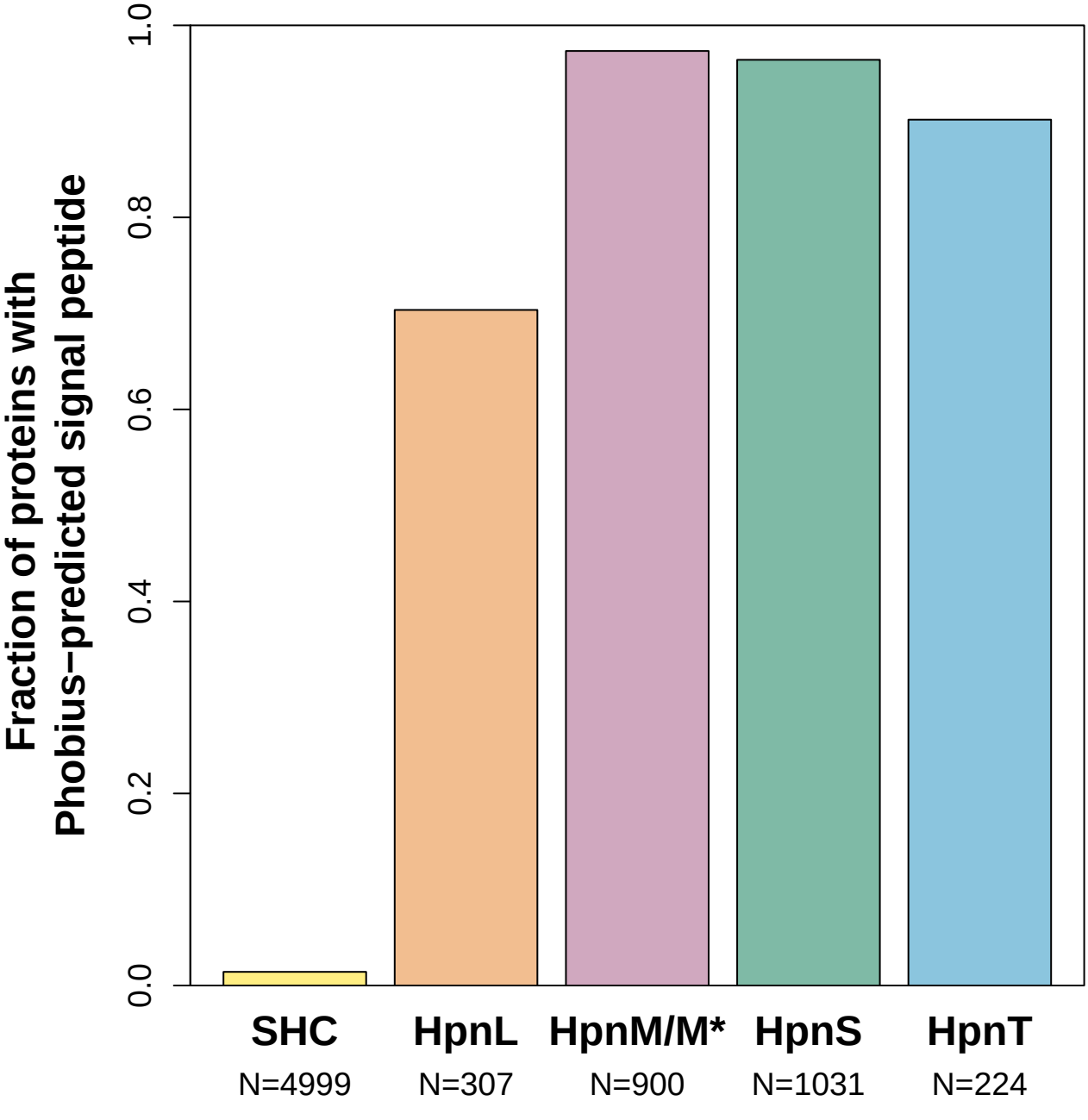

0.53

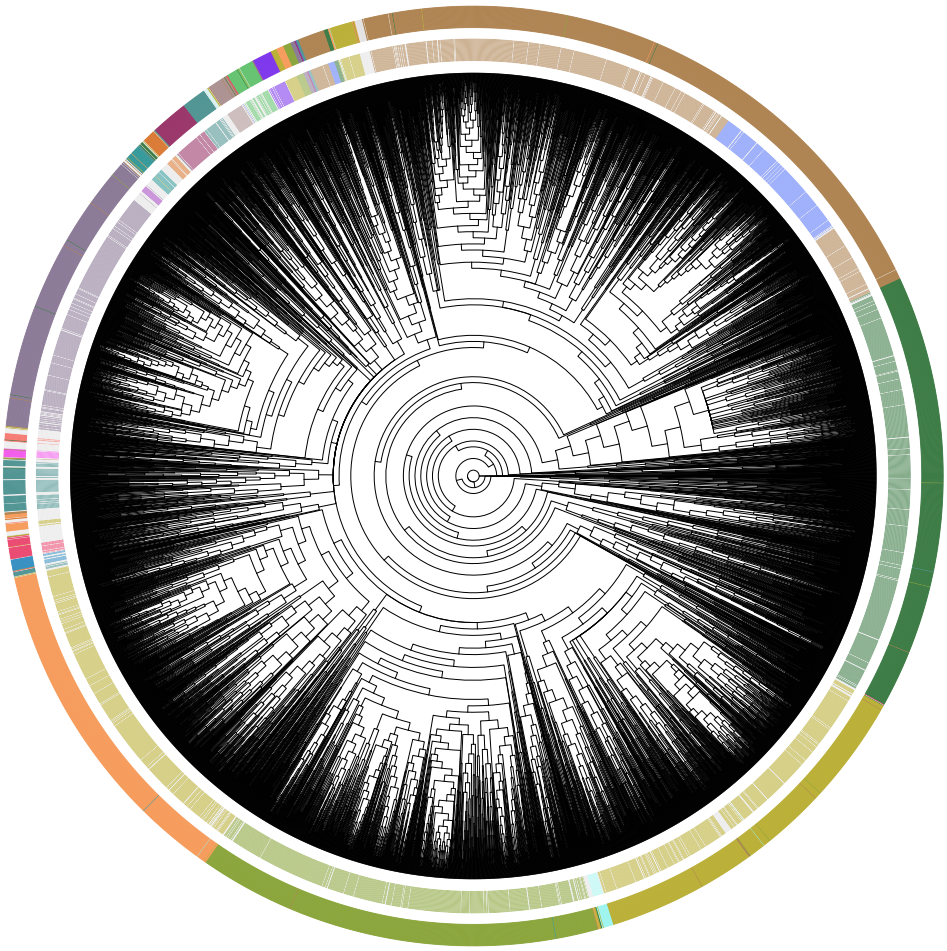

Omnitrophota
Armatimonadota
Elusimicrobiota
OUTGROUP
Verrucomicrobiota
Planctomycetota
Bacillota
Bacteroidota
Chloroflexota
Acidobacteriota
Actinomycetota
Nitrospinota
Nitrospirota
Cyanobacteriota
Myxococcota
Desulfobacterota_B
Desulfobacterota
Gammaproteobacteria
Alphaproteobacteria
Magnetococcia
Zetaproteobacteria
0.52


Hyphomicrobium
Rhizomicrobium
Afipia
Sphingomonas
Magnetospira
Pandoraea
Bradyrhizobium
Nitrobacter
Paraburkholderia
Rhodopseudomonas
Caballeronia
Tardiphaga
Pseudorhodoplanes
Acetobacter
Burkholderia
Novosphingobium
Parvularcula
Salinisphaera
Telluria
Cupriavidus
Trinickia
Azotobacter
Methylomarinum
Methylobacter
Methylomicrobium
Nitrosospira
Pyruvatibacter
Acidithiobacillus
Roseococcus
−15000
−10000
−5000
0
5000
10000
15000
Genomic Position
Hopanoid
Isoprenoid
hpnM
hpnS
hpnT
hpnL
Other


Bacillota_DEFGI
Bacillota_B
Bacillota_C
Bacillota_A
Fusobacteriota
Bacillota
Chloroflexota
Thermotogota
Deinococcota
Acidobacteriota
Desulfobacterota
Desulfobacterota_BCEF
Campylobacterota
Alphaproteobacteria
Gammaproteobacteria
Bacteroidota
Spirochaetota
PVOP Group
Synergistota
Cyanobacteriota
Actinomycetota
Domain Conservation
>80%
>60%
>40%
>20%
>0%
Absent
Rings (Inner to Outer)
C30 Hopanoid Production
SHC (MetaCyc PWY-7072)
SHC (Kegg Ontology K06045)
SHC-C (Pfam pfam13243)
SHC-N (Pfam pfam13249)
SHC (TIGRFAM TIGR01507)


NH4Cl
KNO3
NaNO2
NH2OH
WT+E
Δshc+E
Δshc+C
OD600
time/hours
0.0
0.2
0.4
0.6
0.8
1.2
0
24
48
72
96
120
144
168


Fraction of proteins with
Phobius−predicted signal peptide
0.0
0.2
0.4
0.6
0.8
1.0
SHC
N=4999
HpnL
N=307
HpnM/M*
N=900
HpnS
N=1031
HpnT
N=224



Assigned Phylum

Ordered clockwise, from top:

Pseudomonadota:

- Gammaproteobacteria
- Betaproteobacteria (NCBI)
- Alphaproteobacteria

Bacillota

Fusobacteriota

Actinomycetota

Bacillota A-I (GTDB)

Synergistota

Thermotogota

Thermodesulfobacteriota (NCBI) / Desulfobacterota (GTDB)

Verrucomicrobiota

Planctomycetota

Bacteroidota

Chlorobiota (NCBI)

Spirochaetota

Acidobacteriota

Campylobacterota

Cyanobacteriota

Chloroflexota

Deinococcota