The Phage-shock-protein (PSP) Envelope Stress Response: Discovery of Novel Partners and Evolutionary History

Janani Ravi[1,2]*, Vivek Anantharaman[3], Samuel Zorn Chen[4], Evan Pierce Brenner[1], Pratik Datta[2], L Aravind[3*], Maria Laura Gennaro[2]*.

[1]Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; [2]Public Health Research Institute, Rutgers University, Newark, NJ, USA; [3]National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD, USA; [4]Computer Science Engineering Program, Michigan State University, East Lansing, MI, USA.

*Corresponding authors:

janani.ravi@cuanschutz.edu; aravind@nih.gov; marila.gennaro@rutgers.edu

Running title: *PSP evolution across the tree of life*

Abstract

Bacterial phage shock protein (PSP) systems stabilize the bacterial cell membrane and protect against envelope stress. These systems have been associated with virulence, but despite their critical roles, PSP components are not well-characterized outside proteobacteria. Using comparative genomics and protein sequence-structure-function analyses, we systematically identified and analyzed PSP homologs, phyletic patterns, domain architectures, and gene neighborhoods. This approach underscored the evolutionary significance of the system, revealing that the core PspA gene (Snf7 in ESCRT outside bacteria) was present in the Last Universal Common Ancestor (LUCA), and that this ancestral functionality has since diversified into multiple novel, distinct PSP systems across life. Several novel partners of the PSP system were identified: (i) the Toastrack domain, likely facilitating assembly of sub-membrane stress-sensing and signaling complexes, (ii) the newly-defined HAAS-PadR-like transcriptional regulator pair system, and (iii) multiple independent associations with ATPase, CesT/Tir-like chaperone, and Band-7 domains in proteins thought to mediate sub-membrane dynamics. Our work also uncovered links between the PSP components and other domains, such as novel variants of SHOCT-like domains, suggesting roles in assembling membrane-associated complexes of proteins with disparate biochemical functions. Results are available at https://jravilab.org/psp.

Importance

Phage shock proteins (PSP) are virulence-associated, cell membrane stress-protective systems. They have mostly been characterized in proteobacteria and firmicutes. We now show that PSP systems were present in the Last Universal Common Ancestor, and that homologs have evolved and diversified into newly identified functional contexts. Recognizing the conservation and evolution of PSP systems across bacterial phyla contributes to our understanding of stress response mechanisms in prokaryotes. Moreover, the newly discovered PSP modularity will likely prompt new studies of lineage-specific cell-envelope structures, lifestyles, and adaptation mechanisms. Finally, our results validate use of domain architecture and genetic context for discovery in comparative genomics.

Keywords

Phage shock protein response | PSP | Envelope stress response | Phylogeny and evolution | Domain architectures | Genomic contexts and neighborhoods | Comparative genomics

Introduction

Cell membranes are complex, dynamic structures made of bilayer and non-bilayer lipids and proteins (1). These membrane components engage in critical processes, including membrane biogenesis, cell shape maintenance and division, small molecule transport, inter- and intracellular signaling, maintenance of proton motive force (PMF), and motility through cytoskeletal proteins (2). Cell membranes continuously adapt to external stresses, particularly in unicellular organisms (3–7). Failure to adjust leads to membrane damage and cell death. A specialized suite of lipids and proteins maintain membrane form and function under stress (8, 9), including the multiprotein bacterial Phage Shock Protein (PSP) system, and the archaeo-eukaryotic ESCRT system (10–20).

The origins and divergence of membrane structure and function remain open questions in the very basis of life. Most membrane functions are conserved across the tree of life despite compositional and structural variations (1), which provides insights into evolutionary history and niche specialization. Where functions differ can also be informative. Prior research has validated that evolutionary processes across life are constrained, and the potential evolutionary paths are not limitless (21, 22). Mapping out evolutionary history is, therefore, more attainable by the existence of constrained characteristics – a basis for phenotypic convergence. Membrane maintenance mechanisms obey these same rules, and characterizing membrane homeostasis systems like PSP can provide functional and evolutionary insights across life.

PSP functionality is driven by the peripheral membrane effector protein PspA, and membrane stability phenotypes are restored in *E. coli* Δ*pspA* strains by complementation with the homologous plant protein, Vipp1 (11–16, 23–31). Both proteins contain the coiled-coil structure PspA domain, and both have been shown to form large homooligomeric complexes (16, 17, 19, 20, 32, 33) (**Table**

S1). Bacterial PspA senses and responds to extracellular stress (14, 23, 33–35). Its domain, Pfam PspA_IM30 (inner membrane 30KDa domain) is found in bacteria, photosynthetic eukaryotes, and archaea (15–17, 33, 36, 37) (**Fig. 1A**). Proteins containing PspA_IM30 are known to facilitate membrane fusion (16) and prevent leakage of protons through the membrane (2, 23) in response to a wide range of surface stresses, including changes in membrane potential, *ΔpH* (2, 23) and PMF (38–41), (I)antibiotic stress (7), heat/osmotic shock (42) and mechanical stress via stored curvature elastic stress (13, 43–45). Vipp1 is involved in thylakoid biogenesis, vesicular lipid transport and storage, and membrane shaping (16, 26, 27, 32).

The classical PSP system first discovered in Proteobacteria (46) is not conserved across all bacterial and archaeal clades (12, 33, 46). However, the PSP systems described in Firmicutes and Actinobacteria (3, 11, 14, 15, 35, 36, 47, 48) include clade-specific PSP proteins with distinct domain architectures and conserved functions (*e.g.,* transcriptional regulation, membrane tethering and relocation). Several key questions remain, including how PspA evolved to function with distinct partners across diverse lineages, about the identities of other integral components of the PSP system beyond the classical systems and complete characterization of cognate partners, if any homologs participate in novel functions, and whether common themes emerge from the phyletic patterns of PSP systems.

We set out to evaluate the evolutionary and functional significance of the PSP system through a comprehensive analysis of PspA occurrences, associated genomic contexts, and phyletic spreads across the three superkingdoms or domains of life (49, 50). To perform this, we resolved PSP proteins into their constituent domain architectures and queried each individual domain in our comparative genomics and phylogenetic analyses. This domain-centric approach identified several novel players in the PSP envelope stress response system, of which we described putative function and evolution. We have also made available the entire PSP repertoire — including new members, affiliate proteins, homologs (including paralogs and orthologs), domain architectures, genomic neighborhoods, phyletic spreads, and phylogenetic analyses in all representative species — for researchers through tabular and graphical summaries on our user-friendly, interactive companion web application.

Results and Discussion

PspA is found widely across the tree of life in distinct genomic neighborhoods and with different functional partners (15, 30, 51). Using comparative genomics and a computational molecular evolutionary approach, we therefore sought to comprehensively identify and characterize all homologs of PspA and cognate partner domains across a dataset representative of the tree of life. We discovered several novel conserved, lineage-specific domain architectures and genomic contexts of PSP systems and delineated their phyletic spreads across the tree of life in the process (see *Methods* for details). All our findings (data, results, and visual summaries) are available in an interactive and searchable web application (https://jravilab.org/psp) that can be queried with the protein accession numbers cited inline throughout the *Results* section.

Phyletic spread of classical PSP components across the tree of life

To describe the network of PSP proteins across the tree of life we analyzed the classical PSP components in Proteobacteria, Actinobacteria, and Firmicutes. Among Proteobacteria, we traced the *E. coli psp* operon, which is widely conserved across Gammaproteobacteria (10, 12, 39, 52–55). PspA, PspB, and PspC are each composed of their namesake domains spanning almost their entire protein lengths. The transcriptional regulator PspF is encoded on the opposite strand and shares a central promoter region with the *pspABC* operon in most Gammaproteobacteria, and only with *pspA* in other organisms. PspF is an enhancer-binding protein, with a NtrC-type AAA+ domain fused to a C-terminal Fis-type HTH (helix-turn-helix) DNA-binding domain. In Actinobacteria, the *M. tuberculosis psp* operon (14, 15, 36) includes the namesake PspA protein; a transcription factor, ClgR, unique to Actinobacteria and containing a Cro-like-HTH (cHTH) DNA-binding domain; an integral membrane protein, Rv2743c, distinct from *E. coli* PspC and PspB; and Rv2742c, an uncharacterized protein carrying a "Domain of Unknown Function,'' DUF3046 (14, 15, 36). For consistency, we rename Rv2743c and Rv2742c as PspM and PspN, respectively (15). *B. subtilis,* a member of Firmicutes, contains the *lia* operon. LiaI is a small transmembrane (TM) protein (11, 48, 56), LiaF contains N-terminal TM helices (DUF2157) and a C-terminal globular domain (DUF2154), and LiaG contains an uncharacterized domain DUF4097. We defined two novel domains of the PSP system in these operons: the LiaI-LiaF-TM sensory domain that unifies the four TM domains (4TM) and DUF2157 from LiaI and LiaF, and Toastrack, unifying DUF4097/DUF2154/DUF2807 Pfam models. The *lia* operon is controlled by a two-component system LiaRS (11, 35). LiaR contains a receiver domain, and a winged HTH (wHTH) DNA-binding domain (REC-wHTH), while LiaS is a sensor kinase with a TM region, intracellular HAMP, and a histidine kinase (HISKIN) signaling domain (57). Information on domains in this section is presented in **Text S1.1** and **Table S1**, and graphically in **Fig. 1A** and **1B**.

PspA

From six distinct queries (Proteobacteria, Firmicutes, Actinobacteria, Cyanobacteria, Plants; see *Methods*), homologs of PspA were sought comprehensively. Consistent with prior studies, PspA_IM30 was identified in most bacterial and archaeal lineages (**Fig. 1C**; web app). Within eukaryotes, only the SAR and Archaeplastida (greater plant lineage) carry PspA_Vipp1 homologs

(**Fig. 1C**). Analysis using multi-query sequence searches, we also identified Snf7 proteins from the eukaryotic membrane-remodeling ESCRT systems (14, 18, 58–61) as homologs (**Fig. 1B**). Snf7 was then used for iterative searches, which yielded several pan-eukaryotic and archaeal homologs distinct from Vipp1 (**Fig. 1B** and **1C**; **Table S3**), supporting a considerably wider evolutionary distribution for PspA, corroborated by other recent reports (19, 20).

Toastrack and LiaI-LiaF-TM

Iterative domain-centric searches and multiple sequence alignments (MSAs) revealed that LiaI and LiaG were similar to N-and C-terminal regions of LiaF. We define these two novel domains as **Toastrack** (unifying DUF2154, DUF4097, DUF2807) and **LiaI-LiaF-TM**. Toastrack-containing proteins are characterized by a single-stranded right-handed beta-helix fold with a unique N-terminal 7-stranded region featuring complex intertwining of the beta-helix strands (e.g., PDB: 4QRK, 4OPW). They are pan-bacterial and sporadic in archaea and eukaryotes (**Fig. 1C**). Toastrack is frequently coupled to PspC, LiaI-LiaF-TM, or both (**Fig. 4**; **Table S4**). LiaI-LiaF-TM contains a 4TM region, and is primarily in bacteria, with transfers to Euryarchaeota (**Fig. 1B**). Search methodology is described in **Text S1.1**.

PspB and PspC

PspB, often part of the proteobacterial *psp* operon, is another integral membrane protein implicated in stress sensing and virulence (62–64). PspB has an N-terminal anchoring TM helix followed by an α-helical cytoplasmic globular domain. While PspB is rarely found outside proteobacteria (**Fig. 1C**), we discovered previously unrecognized divergent PspB domains fused to PspC (see section on *PspC architectures)*.

PspC was first identified in the proteobacterial PSP system; it is critical in sensing membrane stress and restoring envelope integrity. Recent studies demonstrate that PspC in some bacterial species may function independently of PspA in response to certain membrane stressors (58, 63, 65, 66). Another integral membrane protein, our analysis showed PspC has two TM helices, the first being a cryptic TM. We observed that the PspC domain is pan-bacterial and also present in a few archaeal clades (**Fig. 1C**). Our analysis showed that PspC has a conserved arginine between two TM helices, the first being a cryptic TM. We also identified several novel domain architectures and genomic contexts of PspC (see section on *PspC architectures*).

PspM and PspN

Domain-centric searches found no discernible homologs for PspMN outside Actinobacteria (**Fig. 1C**; web app), as with prior full protein searches (15, 36, 67). Two TM helices make up the corynebacterial integral membrane partner **PspM** (Rv2743c). Our analyses revealed minimal PspM variation by MSA and a narrow phyletic spread restricted to mycobacteria and corynebacteria (**Fig. 1B**) (15, 36, 67).

The fourth member in the Mycobacterial operon, PspN (Rv2742c), contains a C-terminal DUF3046 and an N-terminal domain we call **PspN_N** (**Table S1**; **Text S1.2**) (15, 36, 67). **DUF3046** is prevalent only in Actinobacteria (**Fig. 1C**). Moreover, the *M. tuberculosis* genome carries a second copy of DUF3046 (Rv2738c), located four genes downstream of *pspN*. We infer that Rv2738c, rather than *pspN*, contains the ancestral copy of DUF3046 (**Text S1.2**). DUF3046 found as part of PspN is likely a duplicate translocated into the PspN open reading frame (ORF) of mycobacteria, especially the *M. tuberculosis* complex. Further, unlike the pan-actinobacterial DUF3046, the N-terminal region, PspN_N, is conserved only in *M. tuberculosis*, with remnants of the coding region existing as potential pseudo-genes or degraded sequences in a few closely related mycobacteria (*e.g., M. avium*).

Evolution of PspA

Since our results support recent findings that eukaryotic Snf7 is a homolog of PspA (19, 20), we used a structure-informed MSA and phylogenetic tree ('Phylogeny' tab, web app; **Fig. 2; Table S1**) to investigate the PspA/Snf7 superfamily and define its evolutionary origins.

Identifying clades carrying PspA and Snf7

In our analyses, all versions of the PspA domain phylogenetically more similar to the 'classical' PspA domain than the Snf7 domains were termed PspA+. We found that most bacterial lineages contain PspA+ clade members (**Fig. 1C**), with a few transfers to archaea. Among eukaryotes, only those containing plastids and two flagella (comprising the SAR/HA supergroup and excavata) have members of this clade (**Fig. 1C**; **Fig. 2**; web app; PDF version of the tree is searchable by the unique leaf identifiers carrying the accession, species, phylum) (59, 60).

We found that the curated set of PspA-like proteins contains a divergent cluster of Snf7 proteins in the ESCRT-III family (Snf7 family on Pfam/Interpro, part of PspA clan; **Table S1**), which we refer to as the Snf7+ clade (**Fig 2**; web app). The ESCRT-III complex is required for endosome-mediated trafficking via multivesicular body formation and sorting and has a predominantly archaeo-eukaryotic phyletic pattern (**Fig. 1C**) (18, 61, 68–71).

Tracing the roots of the PspA/Snf7 superfamily to the last universal common ancestor (LUCA)

We explored the evolution of the PspA-Snf7 superfamily using an MSA of comprehensively selected PspA/Snf7 homologs from distinct clades across the tree of life (49, 50) and different domain architectures (**Fig. 2**). The MSA revealed a unique insertion of heptad repeats in actinobacterial PspA sequences, likely a lineage-specific adaptation. A few cyanobacterial PspA homologs similar to the ancestral plant variant Vipp1 also contained a C-terminal extension (16, 32, 72). A maximum-likelihood phylogenetic tree was generated using this MSA (**Fig. 2**; web app), which showed clear clustering into i) easily distinguishable clades of homologs from Actinobacteria, Firmicutes, and Proteobacteria; ii) the eukaryotic Vipp1 clade grouped with cyanobacterial homologs indicating plastid origins; and iii) Snf7 domain-containing homologs from archaea and eukaryotes forming a well-defined, separate branch. Together, these observations support that the PspA/Snf7 superfamily is ubiquitous across the superkingdoms/domains of life and originated from a PspA-like gene present in the last universal common ancestor (LUCA), as independently corroborated by recent studies (19, 20).

PspA: Novel architectures and neighborhoods

We used our comprehensive, domain-level search to delineate the domain architectures and genomic contexts of the PspA homologs in organisms from diverse clades.

Domain Architectures

Our searches revealed that the underlying domain architectures of most PspA homologs (>98%; ~2500 homologs) show little variation (web app). In most lineages, PspA homologs contain only the characteristic PspA_IM30 domain (**Fig. 3**; **Table S3;** web app). The remaining <2% of homologs contain PspA_IM30 repeats in tandem or fusions with other domains, including NlpC/P60 and a novel PspA-associated (PspAA) domain (**Fig. 3**; **Text S2.1**). The PspA_IM30–NlpC/P60 hydrolase fusion is predicted to catalyze the modification of phosphatidylcholine, likely altering membrane composition (**Fig. 3**; **Table S3**; AFZ52345.1) (73). Similarly, Snf7 showed minimal variation in domain architecture, with occasional fusions (<5%) in eukaryotes, with some potentially attributable to genome annotation errors. Some Actinobacteria have an Snf7 homolog fused to an RND-family lipid/fatty acid transporter, flanked by two genes encoding a *Mycobacterium*-specific TM protein with a C-terminal cysteine-rich domain (74).

Paralogs

We found that genomes encoding multiple copies of PspA do not maintain domain architecture and genomic context among paralogs (**Fig. 3**; **Table S3**; ‘Phylogeny –> Paralog’ tab, web app). In cyanobacteria, dyads/triads of PspA paralogs occur as adjacent, occasionally fused genes (**Fig. 3**; **Table S3**; *e.g.,* BAG06015.1). Consistent with earlier reports (25, 27), we observed that these neighboring PspA genes are part of two distinct clusters of homologs, one resembling the bacterial PspA and the other the eukaryotic (plastid) Vipp1 (**Fig. 3**). Snf7 gene clusters also appear occasionally in archaeal and eukaryotic species (**Table S3**; *e.g.* CBY21170.1). This suggests tandem gene duplication and may relate to the role of these clusters in membrane stabilization as oligomeric complexes, as in the eukaryotic ESCRT systems (75). Further expansion on PspA/Snf7 paralogs and their likely evolution, including inference of gene duplication *vs.* horizontal gene transfer from domain architectures and genomic contexts, is available in the web app. Occasionally, PspA and Snf7 co-occur in conserved gene neighborhoods (**Fig. 3**; **Table S3**) such as in Deltaproteobacteria (AKJ06548.1); in Bacteroidetes (CBH24266.1), these neighborhoods also contain a third gene encoding a coiled-coil protein with a structure reminiscent of the PspA coiledcoils (**Table S3**). These functional considerations are consistent with recent structural and phylogenetic studies indicating that Vipp1, PspA, and Snf7 proteins have arisen from a common ancestor and play similar roles (19, 20).

Novel variations of classical Psp genomic contexts

We next explored the gene-neighborhood themes for each Psp member (**Fig. 3**; **Table S3**). Major themes representing variations of classical Psp operons are summarized below. For the full list of genomic contexts and phyletic spreads, see web app: 'Data,' 'Genomic Contexts' tabs.

Diversity of the pspFABC *operon*

In addition to Proteobacteria, the classic *pspABC* occurs in nitrospirae and some spirochaetes (**Fig. 3**; **Table S3**). There are also variations to this theme (web app): 1) PspC is fused to a divergent C-terminal PspB in addition to a single PspB in the operon (**Fig. 3**; **Table S3**; *e.g.,* ANW39986.1); 2) PspB duplicates appear in the operon (**Fig. 3**; **Table S3**; *e.g.,* AOL22920.1); 3) PspD occurs along with this operon only in Gammaproteobacteria (**Fig. 3**; **Table S3**; *e.g.,* ANW39986.1). Key variations in the transcription regulator PspF (NtrC-AAA and HTH) include: 1) an operon with additional genes for PspB and PspC fusions and Toastrack-containing proteins in a few spirochaetes; 2) operons in Gammaproteobacteria with genes encoding NtrC-like transcription factors with N-terminal ligand-binding ACT and PAS domains (**Fig. 3**; **Table S3**; *e.g.,* AEN65073.1) (76), and a further protein of the DO-GTPase family, predicted to play a role in membrane-related stresses (77, 78), which occasionally feature an additional PspF; and 3) in some Alphaproteobacteria, the PspC in PspFABC has been replaced by multiple Toastrack-containing proteins (**Fig. 3**; **Table S3**; *e.g.,* ANQ40502.1). Thus, classic *pspABC* presents variation in structure and encoded function.

Associations with Vps4 and other classical AAA+-ATPases

The core of an ESCRT complex in archaea is defined by the co-occurrence of one or more copies of Snf7 (*e.g.,* OLS27540.1; [**Table S3**]) and a gene encoding a **VPS4-like AAA+-ATPase** (with an N-terminal MIT and C-terminal oligomerization domain [**Table S2**]) (79). In our analysis, we observed diversity in *vps4* gene neighborhoods (**Text S2.2**). ESCRT complexes are well-studied in eukaryotes, but less so in archaea, where they were originally associated with cell division and named for this role (Cdv system) (80). Crenarchaeota and Thaumarchaeota contain *cdvABC*. *cdvA* has no eukaryotic homolog, while *cdvB* and *cdvC* encode homologs of ESCRT-III proteins and Vps4, respectively. Recent studies have shown that eukaryotes originate within the archaeal clade of Asgardarchaeota (80). Consistent with this, Asgardarchaeota show a conserved gene neighborhood coupling Snf7, a VPS4-like AAA+ ATPase gene, and a gene coding for ESCRT-II wHTH-domain (81) corresponding to the cognate eukaryotic ESCRT complex. These operons may be further extended with additional copies of Snf7 genes and other genes coding for a TM protein and an ABC ATPase. Additionally, examples of transfers of VPS4-like AAA+-ATPase from archaea to several bacterial lineages were observed (**Table S3**; *e.g.,* ACB74714.1). In these cases, *snf7* is displaced by an unrelated gene containing TPR repeat and 6TM, suggesting membrane proximity. The bacterial PspA (**Table S3**; *e.g.* AEY64321.1) may occur with a distinct protein containing two AAA+-ATPase domains – with the N-terminal copy inactive – in some bacterial clades (**Text S2.2**). Both PspA and the membrane-associated Snf7, along with the AAA+-ATPase, may occur in longer operons with other genes encoding an ABC-ATPase, permease, and a solute- and peptidoglycan-binding protein with PBPB-OmpA domains (**Table S3**; *e.g.,* OGG56892.1). These associations between particular bacterial PspA and Snf7 clade members with AAA+-ATPases suggests a role in ATP-dependent membrane remodeling (**Fig. 3**; **Table S3**), comparable to Snf7 and VPS4 in the ESCRT-III-mediated membrane remodeling in the archaeo-eukaryotic lineage (19, 20, 61, 68, 69, 71, 75). The additional transporter components seen in some of these systems suggest a link between this membrane remodeling and solute transport. Here the OmpA domain could interact with peptidoglycan, while the PBPB domain binds the solute trafficked by the linked ABC transporter.

Operons with CesT/Tir-like chaperones and Band-7 domain proteins

We discovered two novel overlapping genomic associations across various bacteria linking PspA with the Band-7 domain and CesT/Tir-like chaperone domain proteins (**Fig. 3**; **Tables S2** and **S3**). Band-7 has been implicated in macromolecular complex assembly, playing a chaperone-like role in binding peptides (82, 83), and in both membrane-localization and membrane dynamics (84). CesT/Tir chaperone domains have been shown to mediate protein-protein interactions in the assembly and dynamics of the Type-III secretion systems of proteobacteria (85). Using profile-profile searches, we identified a previously uncharacterized protein encoded by genes linked to the yjfJ family of *pspA* genes (*e.g.,* ANH61663.1) as a novel member of the CesT/Tir superfamily (**Fig. 3**; **Tables S1** and **S2**; *e.g.,* ANH61662.1). We also observed that the proteobacterial proteins in the neighborhood of *pspA* (**Fig. 3;** *e.g.,* BAB38581.1), typified by YjfI from *E. coli* (DUF2170; **Table S2**; *e.g.,* BAB38580.1), contained a CesT/Tir superfamily domain (86).

The dyad encoding CesT/Tir–PspA forms the core of one class of conserved gene neighborhood (**Table S2**). We observed it as either a standalone unit or within larger operons, such as one group combining CesT/Tir–PspA (**Fig. 3**; **Table S3**; ANH61663.1) with i) a gene encoding a membrane-associated protein with the domain architecture TM+Band-7+coiled-coil+flotillin, ii) a novel AAA+-ATPase fused to N-terminal coiled-coil and β-propeller repeats, and iii) a 3TM domain protein prototyped by *E. coli* YqiJ and *B. subtilis* YuaF (previously implicated in resistance to cell wall-targeting antibiotics) (87, 88). In some Proteobacteria, this operon also contains a phospholipase D (HKD) superfamily hydrolase (**Fig. 3**). Related abbreviated operons coding only for the Band-7 and flotillin-domain-containing protein, the YqiJ/YuaF TM protein, and in some cases, the aforementioned AAA+-ATPase protein is more widely distributed across bacteria and archaea. These might function with PspA homologs encoded elsewhere in the genome.

The second major group incorporates the dyad into a 6–7 gene operon (**Fig. 3**; **Table S3**; *e.g.,* AAN56746.1) featuring a gene encoding a spermine/spermidine-like synthase domain (89) fused to an N-terminal 7TM transporter-like domain (**Fig. 3**; **Tables S2** and **S3**; *e.g.,* AAN56744.1). This operon encodes three additional uncharacterized proteins, including the YjfL family 8TM protein (DUF350), a novel lipoprotein (Ctha_1186 domain), and a β-strand rich protein that we predict to be intracellularly localized (**Fig. 3**; **Tables S3** and **S4;** DUF4178; *e.g.,* AAN56747.1). In some cases, the last gene in this operon codes for a ribbon-helix-helix (RHH) domain transcription factor (**Fig. 3**; **Tables S2** and **S3**).

The third group combines the CesT/Tir–PspA dyad (AMJ95269.1) with polyamine metabolism genes encoding an ATP-grasp peptide ligase (AMJ95273.1) related to the glutathionyl spermidine synthetase (**Fig. 3**; **Table S3**). This association was also recently noticed in a study of AdoMet decarboxylase gene linkages (90). The operon also encodes a potassium channel with intracellular-ligand-sensing, a YjfL family 4TM protein (**Tables S2** and **S3**; DUF350), a metal-chelating lipoprotein (DUF1190), and another uncharacterized protein with predicted enzymatic activity (DUF2491). Alternately, two variants of these operons lack the CesT/Tir–PspA dyad: one where its replacement is a gene encoding a distinct protein occurring as either a secreted or lipoprotein version (DUF4247, *e.g.,* CCP45395.1) and a predicted beta-sandwich enzymatic domain with two highly conserved histidines, as well as a conserved aspartic acid and glutamic acid (DUF2617, *e.g.,* CCP45394.1). In the second variant, the dyad is replaced by a protein containing a Band-7 domain fused to C-terminal 2TM and SHOCT domains (see below). These operons encode two further polyamine metabolism genes, an AdoMet decarboxylase, and a polyamine oxidase.

Lastly, the *Bacillus subtilis ydjFGHL* operon typifies a coupling in Firmicutes of *pspA* (**Table S1**; ANX09535.1) with Band-7 and C-terminal Zinc Ribbon (ZnR) domains in YdjL, and two ZnRs followed by an uncharacterized domain related to YpeB featuring an NTF2-like fold and a TM segment in protein YdjG. The conserved partner of PspA in these systems is the membrane-associated YdjH, which is characterized by a so-called "**TPM phosphatase domain**" followed by a low-complexity segment or a long coiled-coil (**Fig. 3**; **Table S3**) (91). Competing hypotheses suggest that this domain either encodes a generic phosphatase based on its plant homolog (91) or is involved in membrane-associated photosystem II repair and respiratory complex III assembly (92, 93). Thus, TPM association with PspA parallels the associations with the chaperone CesT/Tir and Band-7 domains involved in macromolecular assembly. Our result supports the hypothesis that the CesT/Tir, Band-7, and TPM domains associate with PspA to serve a chaperone-like role in the assembly of membrane-linked protein complexes.

Membrane dynamics with Chaperone-like domains

The contextual association of PspA with AAA+-ATPase, two distinct families of CesT/Tir type chaperones, Band-7 and TPM domains further supports functional roles of PspA homologs in recruiting macromolecular assembly machinery to the inner leaf of the cell membrane. Polyamines like spermine/spermidine have been implicated in membrane stability in a concentration-dependent manner (94). The repeated coupling of polyamine metabolism genes we identified in PspA-CesT/Tir operons implies that a PspA-based system of membrane dynamics monitors polyamine concentration or aminopropylation of other substrates (90) to alter membrane structure and membrane-associated protein complexes. We propose that these systems link extracellular stress sensing with intracellular membrane dynamics based on the frequent coupling with cell-surface lipoproteins. In particular, flotillin domain-containing proteins are likely to be recruited to lipid subdomains with special components (*e.g.,* cardiolipin-rich inner membrane) to interface with PspA (88).

The PspAA domain

Our *pspA* gene neighborhood analysis identified a unique partner in archaeal and bacterial phyla: a protein containing a novel domain with 3 helices and a 4-stranded β-sheet occurring in a two-gene cluster with *pspA* (**Fig. 3**; **Table S3**; *e.g.,* AAM04874.1). This domain often occurs independently but is occasionally fused to an N-terminal PspA in Actinobacteria and Chloroflexi (**Fig. 3**; **Table S3**; *e.g.,* ACU53894.1). This **PspA-associated domain A (PspAA)** (**Fig. 3**; **Table S2**; **Text S1.3**) occurs in a PspA–PspAA dyad in a two-component system (*e.g.*, CAB51252.1), and in another system consisting of a membrane-associated metallopeptidase, another novel domain **PspA-associated domain B (PspAB)** (*e.g.*, AAZ55047.1), and an occasional third gene encoding a SHOCT-like domain (**Fig. 3**; **Tables S2** and **S3**; web app).

PspA with PspM (ClgRPspAMN) or Thioredoxin

Certain *pspA* homologs may occur in Cyanobacteria and Actinobacteria as an operonic dyad with a gene encoding an active thioredoxin domain-containing protein (15). PspA in this dyad is typified by corynebacterial RsmP (*e.g.,* AKK09942.1; **Fig. 3**; **Tables S2** and **S3**; **Text S2.3**), which is predominantly found in rod-shaped Actinobacteria and regulates cell shape via phosphorylation (95). These PspA homologs either occur with thioredoxin (*e.g.,* AKK09942.1) or with a ClgR-like cHTH and PspM genes (*e.g.,* CCP45543.1; **Fig. 3, Fig. S1**, **Table S3**). Corynebacterium spp. contain two distinct PspA paralogs each displaying one of the above contextual associations ('Paralog' tab, web app). The same thioredoxin family occurs with a different PspA clade (typically two copies) in Cyanobacteria (*e.g.,* AFZ14666.1; **Fig. 3**; **Table S3**). The association of the ClgR-PspM and thioredoxin and their mutual exclusion in the operon suggest that they act as alternative regulators of the PspA homologs — repressors in the case of the ClgR, and a potential redox regulator in the case of thioredoxin (**Fig. S1**; **Text S2.3**).

Operons with two-component systems

Multiple two-component systems associate with the PSP system. For instance, the *liaIHGFSR* operon from *B. subtilis* has been studied in the context of lantibiotic resistance (*e.g.*, ANX06812.1; **Fig. 3**; **Table S3**) (11). In a few paenibacilli, genes encoding an additional PspA and PspC with a C-terminal coiled-coil have been inserted into a *liaIHGFSR*-like operon (*e.g.,* APB74393.1; **Fig. 3**; **Table S3**). In Actinobacteria (*e.g.*, CAB51252.1; **Fig. 3**; **Table S3**), the PspA–PspAA dyad occurs with a firmicute-like two-component system.

Classic two-component transcriptional signaling system

We identified two-component signaling system operons linked to either *pspA* or *PspC* in Firmicutes, Actinobacteria, and other clades (**Fig. 3**; web app). In these operons, a membrane-bound histidine kinase (HISKIN) communicates an external stress signal to the response regulator (a receiver domain fused to a HTH domain) protein, triggering a transcriptional response. It is very likely that PspA, LiaI-LiaF-TM, and Toastrack tie into a two-component system regulating membrane integrity and permeability in response to the stress signal. In Actinobacteria, where the HISKIN is fused to PspC, the signal is presumably sensed by PspC. Moreover, when *pspC/Toastrack* operons with two-component systems lack *pspA* in their immediate operonic neighborhood, they might recruit PspA proteins from elsewhere in the genome to express stress response functions.

PspA-free variations of domain architectures and gene-neighborhood

We next investigated domain architectures and genomic contexts involving PSP components that lack PspA, including those with the PspC and Toastrack domains.

PspC domain architectures and gene-neighborhood

PspC is present in most bacteria, some archaeal clades of Euryarchaeota, and a few Asgardarchaeota (web app). Some orthologs of PspC are fused to an N-terminal ZnR domain (*e.g.,* ABC83427.1, **Table S4**), while most other homologs occur either in predicted operons or fusions with multi-TM domains, such as the LiaI-LiaF-TM and PspB (**Fig. 4**; **Table S4**). PspC also fuses to diverse signaling domains such as i) HISKIN from two-component systems (see above); ii) a novel signaling domain we term the HTH-associated α-helical signaling domain (**HAAS)**, which overlaps partly with the Pfam model for DUF1700 (**Table S2**), and iii) the Toastrack domain (**Table S1**). Moreover, Actinobacteria contain a Lia-like system without PspA. The HISKIN domain in these systems is fused to N-terminal PspC and LiaF-LiaI-TM domains (*e.g.,* AIJ13865.1) and accompanied by a REC–HTH-containing protein (**Fig. 5**; **Table S4**). This core shares a promoter region with a second gene on the opposite strand containing a PspC domain that is fused to additional TMs and, in some cases, a Toastrack domain (*e.g.,* AIJ13866.1; **Fig. 5**; **Table S4**). These associations strongly imply that PspC acts as a sensor domain to detect and communicate a membrane-proximal signal.

Contextual associations of the Toastrack Domain

The **Toastrack** domain repeatedly emerges as a partner to the PSP system. As noted, genes encoding Toastrack-containing protein occur concurrently with *pspA* in several conserved neighborhoods (**Fig. 4**; **Table S4**) and are fused to *pspC* in *pspA*-free settings. We found that the Toastrack domains tend to be C-terminal to various TM domains such as PspC, LiaI-LiaF-TM, and other multi-TM/single-TM domains unique to these systems (**Fig 4**; **Text S2.4**). These fusions, along with the lack of TM or extracytoplasmic regions within Toastrack (**Table S1**; **Figs. 1** and **4**), strongly suggest that Toastrack is an intracellular domain. Further, the N-terminal regions of several Toastrack architectures contain at least two divergent homologs of the bi-helical SHOCT domain (*e.g.,* CAB15517.1; **Figs. 4** and **5**; **Tables S1** and **S4**), which we call **SHOCT-like** domains, which are partly detected by the Pfam **DUF1707** model (**Text S2.4**) (96). We observed these SHOCT-like domains at both N- and C-termini of various globular domains, including Toastrack. The globular domains fused to the SHOCT and related domains identified in this work may occur independently of them in other contexts. In most cases, the domains are neither fused to signal peptides nor to TM segments to indicate a potential extra-cellular location. Moreover, their known/predicted activities are meaningful only in intracellular contexts. Therefore, we predict that this SHOCT-like domain may play a role in anchoring disparate domains to the inner leaf of the membrane.

The DNA-binding domain **LytTR** may also be encoded by operons carrying Toastrack fused to LiaI-LiaF-TM in Firmicutes (*e.g.,* CAL82154.1; **Fig. 5**; **Table S4**). The Toastrack-containing protein (*e.g.,* CAB15517.1) occurs in an operon with a *pspC* homolog (CAB15516.1) and has been implicated in the activation of the membrane-associated *liaFSR* operon (**Table S1**) and protection against membrane permeabilization (97–99). We also identified *pspA*-free variants of the classic *lia* operon with Toastrack and a two-component system, typified by *vraT* of *Staphylococcus aureus*

(ABD31150.1; **Fig. 5**; **Table S4**). These systems carry LiaI-LiaF-TM and Toastrack with a two-component system (*vraSR*) and are involved in methicillin and cell wall-targeting antibiotic resistance (3, 100–102). A similar organization of LiaI-LiaF-TM and Toastrack containing a protein with a two-component system is found in a few bacterial species (*e.g.,* among Proteobacteria, ABD83157.1, Ignavibacteriae, AFH48155.1; **Table S4**).

We discovered several other conserved gene neighborhoods centered on Toastrack-containing genes likely to define novel functional analogs of the PSP system with potential roles in membrane-linked stress response. One of these, a four-gene context found across diverse bacterial lineages, includes a previously uncharacterized protein similar to the Pfam model DUF2089. The Pfam model **DUF2089** can be divided into an N-terminal ZnR, central HTH, and C-terminal SHOCT-like domains (ADE70705.1), variants of which appear with LiaI-LiaF-TM, B-box, or PspC domains (**Figs. 4** and **5**; **Tables S2** and **S4**; **Text S2.4**). Because of these shared operonic associations, these Toastrack contexts may function similarly to the classical *lia* operon in transducing membrane-associated signals to a transcriptional output affecting a wide range of genes, including through an operonically linked sigma factor.

Similarly, we observed operons coupling a membrane-anchored Toastrack-containing protein with an ABC-ATPase, a permease subunit, a **GNTR-HTH** transcription factor with distinct C-terminal α-helical domain (**Fig. 5**; **Tables S2** and **S4**; **Text S2.4**), and a DUF2884 membrane-associated lipoprotein (*e.g.,* AJI31452.1). The latter might function as an extracellular solute-binding partner for the ABC-ATPase and permease components. We found a similar operon in Actinobacteria with the ribbon-helix-helix (**RHH**) domain transcription factor replacing **GNTR-HTH** (**Table S2**; **Text S2.4**). In some Actinobacteria, the Toastrack domain is fused to a SHOCT-like domain or occurs in a transport operon, likely coupling Toastrack to transcriptional regulation, membrane-proximal signal sensing, and transport (**Fig. 5**; **Table S4**).

Based on these observations, we propose that the two extended β-sheets of the Toastrack domain (formed by the β-helical repeats) facilitate the assembly of sub-membrane signaling complexes by providing two expansive surfaces amenable to protein-protein interactions near the membrane. Protein docking to the Toastrack domain may also activate transcription via fused or associated HTH or LytTR domains, associated two-component systems, or the HAAS-HTH system.

The HAAS-PadR-like wHTH systems

We discovered a novel signaling system that frequently co-occurs with PspC and Toastrack domains in gene neighborhoods that do not contain *pspA* (**Figs. 4** and **5**). This system consists of two components, the HAAS domain and the PadR-like wHTH (**Table S2**). The HAAS domain is partly similar to the Yip1 protein family (Pfam models DUF1700 and DUF1129; Pfam clan: Yip1). An HHpred search further unifies this model with the Pfam profile DUF1048 (PDB: 2O3L). We therefore refer to them collectively as the **HAAS** superfamily. The core HAAS fold has three consecutive α-helices, with the third helix displaying an unusual kink corresponding to a conserved GxP motif, predicted to form part of a conserved groove that mediates protein-protein interactions. The HAAS domain always occurs in gene neighborhoods coupled with a protein containing a standalone PadR-like wHTH DNA-binding domain (**Fig. 5**). This co-migration suggests that HAAS transduces a signal from domains fused to this wHTH transcription factor.

When coupled with PspC, the HAAS domain shows three broad themes: 1) it is found as part of a multidomain TM protein with additional PspC, LiaI-LiaF-TM, and Toastrack domains; 2) it is fused directly to a Toastrack domain in a gene neighborhood that also codes for a PspC domain protein, occasionally fused to other TM domains; 3) it constitutes a part of a TM protein fused to conserved multi-TM domains other than the LiaI-LiaF-TM or PspC (*e.g.,* VanZ) (**Figs. 4** and **5**; **Table S4**). These gene neighborhoods code for standalone *pspC* genes. Additionally, the HAAS domains occur in PspC-independent contexts, but are typically coupled to the N-terminus of other multi-TM or intracellular sensory domains. We found conserved fusions to at least 5 distinct multi-TM domains (**Figs. 4** and **5**; **Table S4**), such as 1) the FtsX-like TM domains with extracellular solute-binding MacB/PCD domains (*e.g.,* ACO32024.1); 2) FtsW/RodA/SpoVE family of TM domains (*e.g.,* CAC98500.1) (103); 3) various uncharacterized 6TM, 4TM, and 2TM domains. In addition to Toastrack, the HAAS domain may combine with other intracellular domains such as the pentapeptide repeat domains mostly in Firmicutes (*e.g.,* AOH56696.1; **Fig. 5**; **Table S4**). The organisms in which these operons occur have PSP components elsewhere in the genome. However, these proteins might be recruited to the stress response system, as demonstrated by the effects of PadR-like wHTH deletion studies in the *Listeria* Lia systems (103).

In bacteria lacking the typical *lia* operon (carrying the LiaRS-like two-component system), we noted that the HAAS–PadR-like-wHTH dyad replaces the histidine kinase-receiver domain transcription factor pair. Therefore, we propose that the HAAS-PadR dyad constitutes an alternative to the classical two-component systems. Here, the HAAS domain is proposed to undergo a conformational change in response to membrane-linked or intracellular signals detected by the sensor domains (*e.g.,* LiaI-LiaF-TM or PspC) that are fused to or co-occur with the dyad. This conformational change is likely transmitted to the associated PadR-like wHTH via the groove of the conserved GxP motif; this change might lead to transcription factor release and a downstream transcriptional response.

A unified view of PSP partners and evolution

We reconcile our findings on PSP-related genomic contexts and partner domains within the framework of the tree of life to assemble the PSP puzzle. The above results include ~20,000 homologs of key PSP proteins, ~200 novel domain architectures (minimum 2 species), and ~500 genomic contexts (minimum 5 species) (web app). To visualize the broader context of PSP systems across the superkingdoms, we first built a network of domain architectures for the extended PSP system (**Fig. 6A**), with domains as nodes and co-occurrence within a protein as edges. This network consolidates the PSP system with known and novel connections, including those between PspA and PSP partner domains. Next, we summarized the phyletic spread and prevalence of all PSP members and partners in a single heatmap (**Fig. 6B**). This view indicates that, of the PSP components, (i) only PspA/Snf7 is present in all eukaryotes, whereas Toastrack has been transferred from bacteria to certain eukaryotes (*e.g.,* human FAM185A with orthologs in animals and other eukaryotes) ; (ii) PspC and PspA/Snf7 are present in most archaeal lineages, (iii) occasional transfers of Toastrack, LiaI-LiaF-TM, and SHOCT domains are observed from bacteria to Euryarchaeota; (iii) in bacteria, domains such as Toastrack, PspC, PspA, and LiaI-LiaF-TM tend to co-migrate. Finally, using an ‘upset plot,’ we unified: i) the relative occurrence of domain architectures from PSP members and their most common neighbors, ii) combinations of these domain architectures, and iii) their frequencies (**Fig. 6C**). These consolidated visualizations can be explored in our web app. For readers who wish to explore any of these proteins in more depth, the accession numbers from this dataset can also be used for input in our MolEvolvR protein analysis web app (https://jravilab.org/molevolvr) (104).

Conclusion

Recent work with the bacterial PSP stress response system suggests that while function is maintained across phyla, the proteins, regulatory mechanisms, and membrane stress response mechanisms vary widely among lineages (12, 15, 26, 36). These variations reflect the utilization of a common peripheral inner membrane protein in lineage-specific envelope dynamics and stress responses. In this study, we systematically analyzed PSP systems, their phyletic spread, and evolution (**Figs. 1** and **2**), discovering several novel sequence and structural features (*e.g.,* domain architectures) and conserved gene-neighborhoods of PSP components. We identified both classical and newly-predicted partners (**Figs. 2** and **5**) and established their phyletic distributions across the tree of life (**Figs. 2** and **6**). We also established that PspA/Snf7 traces to the LUCA (**Fig. 2**). Indeed, it has been independently shown that the PspA/Vipp1 ESCRT system dates back to the LUCA (19, 20), further strengthening the evolutionary significance of our finding. This confirms that the LUCA possessed a membrane whose curvature and dynamics were mediated by an ancestral PspA/Snf7-like coiled-coil protein assembled into polymeric structures adjacent to the inner leaf of the membrane. This hypothesis is in line with earlier work showing that LUCA contained some signal-recognition particle GTPases (105). These proteins are unequivocal indicators of the presence of a secretory apparatus translocating proteins across lipid bilayers. Hence, as a corollary they are also indicators of the presence of a membrane. Similarly, the extant components of membrane-linked bioenergetics (F1/F0 ATPases or V-Type ATPases) are also indicative of a membrane in the LUCA (106, 107). Thus, we infer that this ancient machinery was robust to the subsequent lineage-specific changes in membrane chemistry, producing different membrane types (*e.g.,* ether-linked vs. ester-linked) in the archaeal, bacterial, and eukaryotic lineages.

Our novel PSP findings make several new functional predictions: 1) Based on its occurrence with membrane partners (*e.g.,* PspA, LiaI-LiaF-TM, PspC, HAAS, SHOCT-like domains) and transcription regulators or two-component systems (*e.g.,* LiaRS, PadR-like-wHTH), we propose that Toastrack is a membrane-proximal intracellular domain serving as a sub-membrane signaling complex assembly platform; 2) The association of PspA homologs with distinct AAA+-chaperones and a non-ATPase chaperone (*e.g.,* CesT/Tir) or predicted chaperone-like proteins (*e.g.,* Band-7, TPM) suggest that these complexes are involved in the assembly of sub- and trans-membrane complexes in response to specific envelope stresses. These observations imply a general role for analogs of ESCRT-like complexes in these bacterial systems; 3) The association of PspA with various transporters and polyamine metabolism systems indicates that the regulation of membrane structure by PspA is associated with changes in solute concentrations and covalent modifications that may affect membrane stability; 4) The discovery of several new versions of the SHOCT-like domain suggests that these domains mediate membrane localization of disparate catalytic and non-catalytic activities, which may subsequently interface with the PSP system; 5) A diverse array of sub-membrane (*e.g.,* Toastrack), TM, and cell-surface domains (*e.g.,* PspC, LiaI-LiaF-TM) co-occur with two-component or HAAS-PadR-like-wHTH systems, hinting at alternatives to conventional two-component signaling that interface with the PSP system in responding to membrane stress.

Further research into these newly described systems will yield broadly applicable mechanistic insights, given their conservation across life. For example, SHOCT domains are understudied since they are thought to be absent in model organisms (96). However, our approach identified divergent homologs in the model organism *B. subtilis* (CAB15517.1) that may prompt new studies. The extensive findings from applying gene neighborhood and domain architecture searches support their viability, and such methods help reduce the space of annotated hypothetical proteins without classified domains and add functional clues. Finally, we provide another pillar of support to recent conclusions that PspA/Snf7 homologs were already present in the last universal common ancestor (19, 20), and deepen our understanding of the origins and diversification of all life. All the findings (data, results, and visual summaries) from this work are found in an interactive and queryable companion web application available at https://jravilab.org/psp.

Methods

Query and subject selection

We queried classical PSP members — PspA (eight representatives across Proteobacteria, Actinobacteria, Firmicutes (two paralogs), Cyanobacteria, and Archaea); PspM (Rv2743c) and PspN (Rv2742c) (from *M. tuberculosis*); PspB and PspC (from *E. coli*); LiaI, LiaG, and LiaF (from *B. subtilis*) — against all sequenced genomes across the tree of life. Transcription factors and two-component systems, including PspF (from *E. coli*), ClgR (from *M. tuberculosis*), LiaRS (from *B. subtilis*), are not part of this detailed analyses or the web app, due to their prevalence across the bacterial kingdom. We ran homology searches for the other representative PSP components against custom databases of ~6500 completed representative or reference genomes with taxonomic lineages or the NCBI non-redundant database (**Tables S1** and **S2**) (108, 109). All PSP homologs are listed in the web app, with representatives in **Tables S3** and **S4**. We obtained the phyletic order (sequence) from NCBI taxonomy and BV-BRC (108–112).

Identification and characterization of protein homologs

We analyzed the classical PSP proteins from the best-studied operons *pspF||ABCDE-G* (from *E. coli*), *liaIHGFSR* (from *B. subtilis*), and *clgRpspAMN* (from *M. tuberculosis*) and their phyletic spread to locate and stratify all PSP homologs. We first resolved these PSP proteins into their constituent domains and used each individual domain as a query for all our phylogenetic analyses to ensure an exhaustive search and identification of all related PSP proteins, including remote homologs. This approach allows us to find homologies beyond those from full-length proteins only. We then performed homology searches for each constituent domain across the tree of life (approximately 6500 representative or reference genomes). We used a combination of protein domain and orthology databases, iterative homology searches, and multiple sequence alignments to detect domains, signal peptides, and transmembrane (TM) regions to help construct the domain architecture of each query PSP protein. The ubiquitous presence of transcription factor HTH domains and histidine kinase-receiver domain two-component systems in bacterial phyla precluded dedicated searches with these domains; instead, we noted their occurrence in predicted operons alongside core PSP genes to identify transcriptional regulation partners.

We ensured the identification of a comprehensive set of homologs (close and remote) for each queried protein via iterative searches using PSI-BLAST (113) with sequences of both full-length proteins and the corresponding constituent domains. More distant relationships were determined using profile-profile searches with the HHpred program (114). Protein searches were conducted using homologous copies from multiple species as starting points. We aggregated search results and recorded the numbers of homologs per species and genomes carrying each of the query proteins (web app). These proteins were clustered into orthologous families using the similarity-based clustering program BLASTCLUST (115). SignalP, TMHMM, Phobius, JPred, Pfam, and custom profile databases (116–121) were used to identify signal peptides, TM regions, known domains, and the secondary protein structures in every genome. Homolog information, including domain architectures, is available in the web app (‘Data,’ ‘Domain architectures’ tabs).

Neighborhood search

Bacterial gene neighborhoods (±7 genes flanking each protein homolog) were retrieved using a Perl script from GenBank (108, 109). Gene orientation, domains, and secondary structures of the neighboring proteins were characterized using the same methods applied to query homologs above. Genomic contexts are available in the web app ('Genomic contexts' tab). We note that eukaryotic components (*e.g.,* Snf7) appearing alone are an artifact of the genomic context analysis being restricted (and relevant) to bacteria and archaea (see web app).

Phylogenetic analysis

Multiple sequence alignment (MSA) of the identified homologs was performed using Kalign (122) and MUSCLE (123) with default parameters. Domain/gene trees were constructed with FastTree 2.1 with default parameters (124) and visualized and labels colored in FigTree v1.4.4 (125). In parallel, the PspA/Snf7 MSA was also passed to ModelTest-NG, which reported the best performing substitution model (LG+I+G4+F) (126, 127), and then RAxML-NG-MPI (v1.2.0) (128) on CU's Alpine HPC cluster and a maximum likelihood phylogenetic tree constructed (--all, –model LG+I+G4+F, –tree rand{25} pars{25}, –bs-trees autoMRE, –threads 60). Bootstrapping was performed until RAxML-NG assessed statistical convergence at 480 replicates (**Fig. S2**). Phylogenetic trees are available in the web app ('Phylogeny' tab).

Network reconstruction

The PSP proximal neighborhood network was constructed on the domain architectures and genomic contexts of PspA and its cognate partner proteins (**Tables S1** and **S2**). The nodes represented the domains, and edges indicated a shared neighborhood (domains of the same protein or neighboring proteins). Proximity networks are available in the web app ('Domain architectures' tab).

Web application

We built our interactive and searchable web application using R Shiny (129). All data analyses and visualizations were carried out using R or RStudio (130, 131) with R packages (132–144). Data and results from our study are available on the web application that serves as a companion data summarization and visualization dashboard to this manuscript (https://jravilab.org/psp). The web app can be queried with the protein accession numbers, domain names, or species/lineages cited inline throughout the Results section. While not used directly in this analysis, our MolEvolvR protein analysis web app can be used with the accession numbers of any protein we describe for readers who would like to further explore a protein's domain architectures and phylogenetic distribution across life (https://jravilab.org/molevolvr) (104).

Figures

Figure 1. The phyletic spread of classical PSP members across all major lineages.

A. The three classical PSP systems in *E. coli* (*pspF||pspABC*), *M. tuberculosis* (*clgRpspAMN*), and *B. subtilis* (*liaIHGFSR*) are shown. Proteins across three bacterial PSP systems are labeled and colors uniformly (across all figures that show domain architectures) indicate the nature of the protein; ***black***, PspA homolog (*e.g.,* PspA, LiaH); ***teal***, transcription factor/response regulator (*e.g.,* PspF, LiaR, MprA, ClgR) along with partner histidine kinases in ***olive green*** (*e.g.,* LiaS, MprB part of two-component systems LiaRS, MprAB); ***orange/yellow***, transmembrane protein (*e.g.,* PspB, PspC, LiaF, LiaI, PspM) embedded in the dashed blue membrane. Arrows represent interaction or activation, or inhibitory feedback in the case of LiaF and LiaS. A simple operon map is shown for each system at the bottom, but is expanded upon in panel **B**. **B. Domain architectures of the**

classical PSP operons in *E. coli*, *M. tuberculosis*, and *B. subtilis*. Domains are denoted by rectangular segments inside the block arrow representing each protein labeled below the arrow (*e.g.*, the protein LiaH contains only the PspA domain). The direction of the arrow indicates the direction of transcription. See *Results* for new domain definitions. **C. Phyletic spreads of PSP proteins.** Sunburst plots are shown to indicate the lineage distributions (as a fraction of the total number of homologs recorded across lineages) for the homologs of 'classical' domains/protein families of interest: PspA, Snf7, PspB, PspC, PspM, PspN (and DUF3046), LiaI-LiaF-TM, Toastrack. In each plot corresponding to a particular protein, the inner ring corresponds to the proportion of its homologs present in superkingdoms/domains of life, Bacteria, Archaea, and Eukaryota. The outer ring depicts the distribution of the homologs among key phyla. Interactive sunburst plots for each of the PSP proteins are available in the web app. The colors for each lineage (outer and inner rings) are shown separately for **C**.

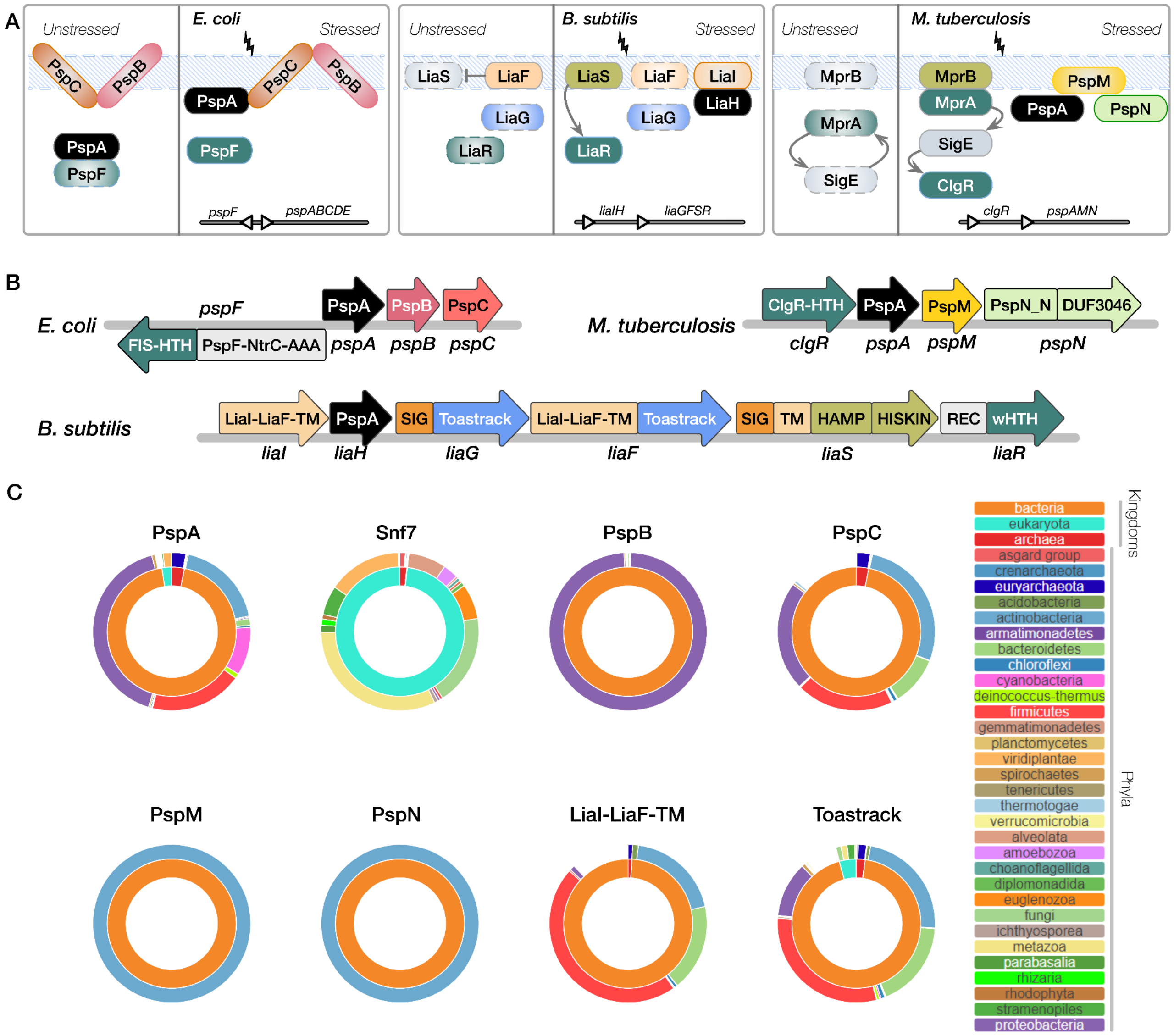

A
Unstressed
E. coli
Stressed
PspC
PspB
PspA
PspF
pspF
pspABCDE
Unstressed
B. subtilis
Stressed
LiaS
LiaF
LiaG
LiaR
LiaI
LiaH
lialH
liaGFSR
Unstressed
M. tuberculosis
Stressed
MprB
MprA
SigE
ClgR
PspM
PspN
clgR
pspAMN
B
E. coli
pspF
FIS-HTH
PspF-NtrC-AAA
pspA
pspB
pspC
M. tuberculosis
ClgR-HTH
PspA
PspM
PspN_N
DUF3046
clgR
pspA
pspM
pspN
B. subtilis
LiaI-LiaF-TM
PspA
SIG
Toastrack
LiaI-LiaF-TM
Toastrack
SIG
TM
HAMP
HISKIN
REC
wHTH
liaI
liaH
liaG
liaF
liaS
liaR
C
PspA
Snf7
PspB
PspC
PspM
PspN
LiaI-LiaF-TM
Toastrack
Kingdoms
Phyla
bacteria
eukaryota
archaea
asgard group
crenarchaeota
euryarchaeota
acidobacteria
actinobacteria
armatimonadetes
bacteroidetes
chloroflexi
cyanobacteria
deinococcus-thermus
firmicutes
gemmatimonadetes
planctomycetes
viridiplantae
spirochaetes
tenericutes
thermotogae
verrucomicrobia
alveolata
amoebozoa
choanoflagellida
diplomonadida
euglenozoa
fungi
ichthyosporea
metazoa
parabasalia
rhizaria
rhodophyta
stramenopiles
proteobacteria

Figure 2. PspA/Snf7 homologs across the tree of life.

Phylogenetic tree of PspA homologs across the tree of life. The phylogenetic tree was constructed using FastTree and visualized with FigTree (parallel RAxML-NG generated PspA tree with confidence values is shown in **Fig. S2**) from a multiple sequence alignment (using kalign3) of representative PspA homologs across all major superkingdoms and phyla (see *Methods*; web app). The primary takehome from the tree is not the 100s of leaves/labels but the key lineages that are labeled next to distinct clusters of similar PspA proteins; leaf colors match the lineage labels to highlight this import. The insets show the 3D structures for PspA (4WHE) and Snf7 (5FD7) from the Protein Data Bank. The tree leaves are labeled by lineage, species, and accession numbers. A text-searchable, scaled vector graphics version of this tree is available through the web app under the Phylogeny tab (leaf labels with aforementioned lineage, species, accession can query this high-resolution downloadable PDF/SVG version of the tree). Representatives for PspA/Snf7 homologs are also available in **Table S3**. As a gene tree, most homologs do cluster among evolutionarily similar species, though evidence of widespread horizontal gene transfer is also evident. Two clusters of eukaryotic genes appear: Snf7 in Eukaryota and Asgardarchaeota groups, mirroring recent research suggesting eukaryotic life arose from this clade; and Vipp1 among Eukaryota and Cyanobacteria group, again reflecting the shared origin of the protein in plants from Cyanobacteria.

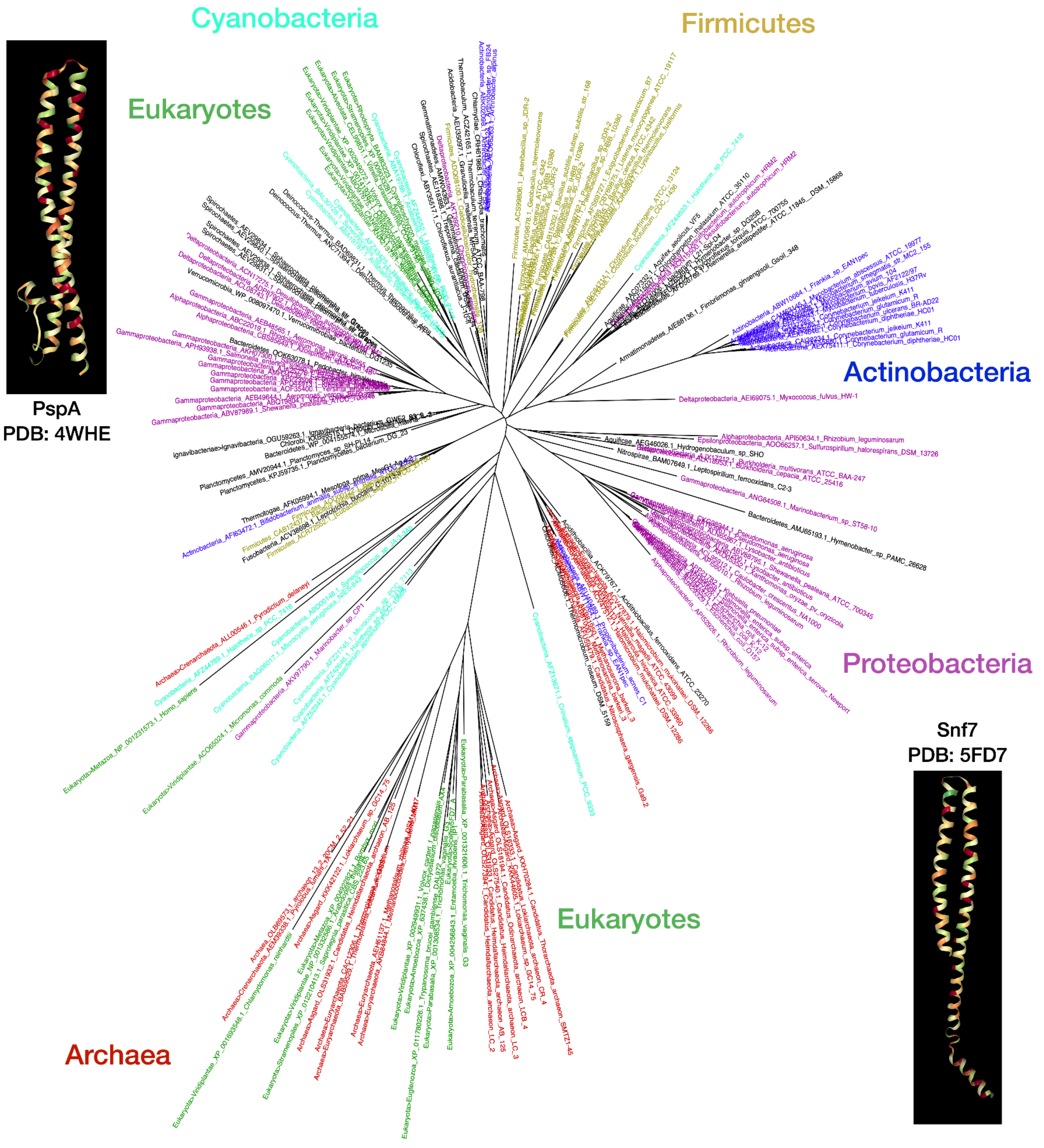

Cyanobacteria
Firmicutes
Eukaryotes
PspA
PDB: 4WHE
Actinobacteria
Proteobacteria
Snf7
PDB: 5FD7
Eukaryotes
Archaea

Figure 3. PspA domain architectures and genomic contexts

The first row contains domain architectures of the most prevalent homologs of PspA (as in **1A**; indicated in ***black*** throughout). The remaining rows show the predominant genomic contexts of PspA homologs across multiple bacterial and archaeal lineages (identified by neighborhood searches ±7 genes flanking each homolog; see *Methods*). Representative neighborhoods with archetypal lineages and archetypal example proteins (with accession numbers and species) are shown. The PspA contexts are grouped by neighboring domains such as (a) PspF, PspB/PspC; (b) PspAA, PspAB; (c) ClgR, PspM/PspN, Thioredoxin (**Fig. S1**); (d) chaperones such as Band-7 proteins like flotillin, CesT_Tir, TPM_phosphatase, ZnR, SpermGS-ATPgrasp, and Spermine synthase (**Table S2**); (e) two-component systems such as the Lia system and Toastrack, and other novel genomic contexts; and (f) cyanobacterial variations. Key: *rectangles*, domains; *arrowheads*, the direction of transcription; *domains enclosed by dotted lines*, absent in the genomic contexts in certain species; *white cross*, substitution with protein(s) mentioned just below; *white triangle*, insertion with one or more proteins; *two vertical lines* '||', indicate a change in the direction of transcription; *small black boxes,* domain repeats co-occurring within a protein/context. Archetypal accession numbers and species are provided mostly on the left. Archetypal lineages are indicated in grey on the right for each of the domain architectures and genomic contexts. Different domains are indicated by the same color scheme as in **Figure 1**. Also, similar domains are given similar hues. For example, membrane proteins are in different shades of orange (SIG, predicted signal peptide, dark orange, PspC, orange, other transmembrane domain, light orange); transcription factors/regulators (including HTH, helix-turn-helix domain) are in teal; DUFs, Domains of Unknown Function, and other domains are in grey. Further details of the domain architectures and gene neighborhoods shown are described in the text and **Table S3**, and the full list of PspA homologs, their domain architectures, genomic contexts, and lineages are shown in the web app (under the ‘Data,’ ‘Domain architectures,’ and ‘Genomic contexts’ tabs).

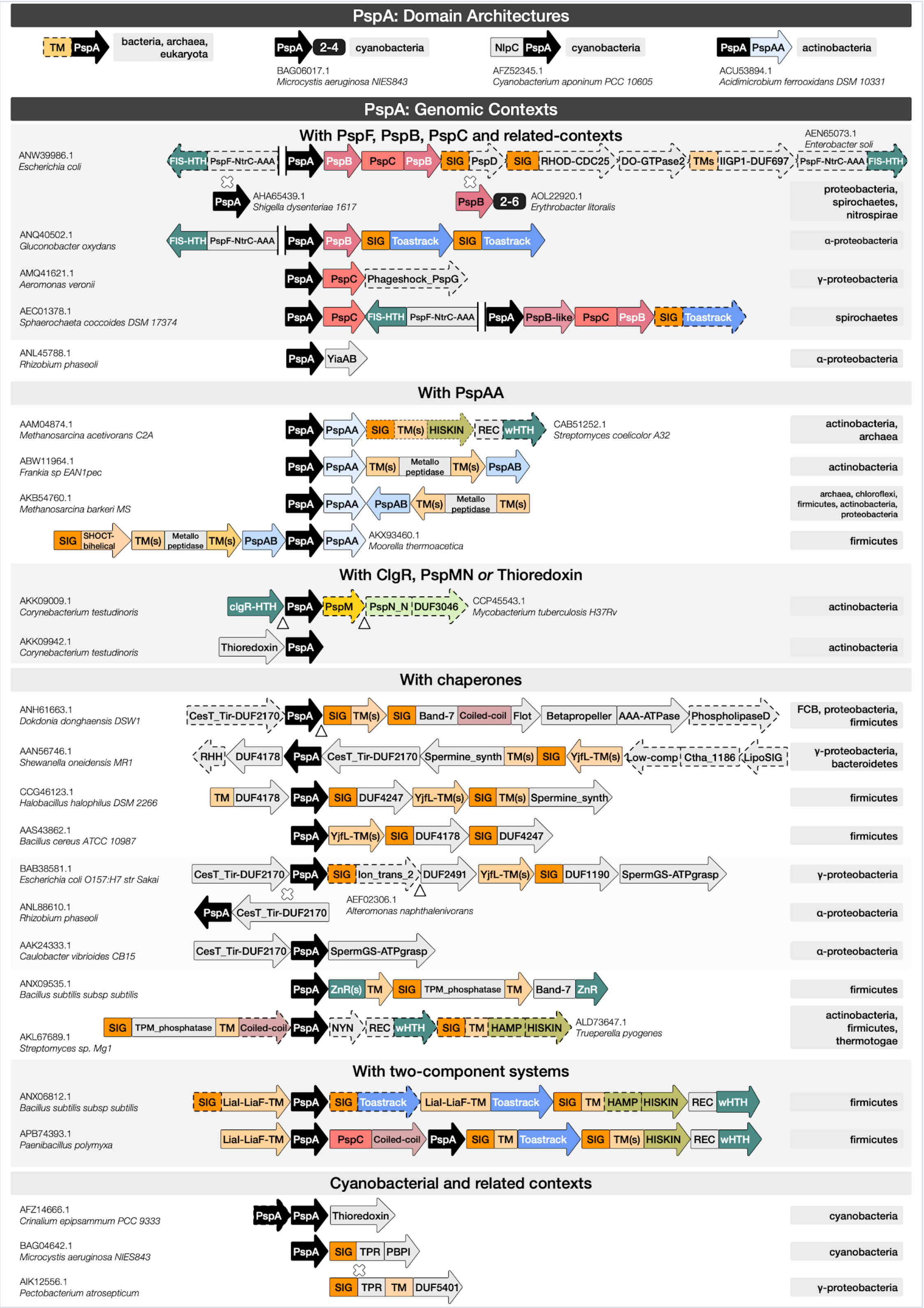

PspA: Domain Architectures
TM PspA
bacteria, archaea, eukaryota
PspA 2-4
cyanobacteria
BAG06017.1
Microcystis aeruginosa NIES843
NlpC PspA
cyanobacteria
AFZ52345.1
Cyanobacterium aponinum PCC 10605
PspA PspAA
actinobacteria
ACU53894.1
Acidimicrobium ferrooxidans DSM 10331
PspA: Genomic Contexts
With PspF, PspB, PspC and related-contexts
ANW39986.1
Escherichia coli
FIS-HTH PspF-NtrC-AAA
PspA PspB PspC PspB SIG PspD SIG RHOD-CDC25 DO-GTPase2 TMs IIGP1-DUF697 PspF-NtrC-AAA FIS-HTH
AEN65073.1
Enterobacter soli
PspA
AHA65439.1
Shigella dysenteriae 1617
PspB 2-6
AOL22920.1
Erythrobacter litoralis
proteobacteria, spirochaetes, nitrospirae
ANQ40502.1
Gluconobacter oxydans
FIS-HTH PspF-NtrC-AAA PspA PspB SIG Toastrack SIG Toastrack
α-proteobacteria
AMQ41621.1
Aeromonas veronii
PspA PspC Phageshock_PspG
γ-proteobacteria
AEC01378.1
Sphaerochaeta coccoides DSM 17374
PspA PspC FIS-HTH PspF-NtrC-AAA PspA PspB-like PspC PspB SIG Toastrack
spirochaetes
ANL45788.1
Rhizobium phaseoli
PspA YiaAB
α-proteobacteria
With PspAA
AAM04874.1
Methanosarcina acetivorans C2A
PspA PspAA SIG TM(s) HISKIN REC wHTH
CAB51252.1
Streptomyces coelicolor A32
actinobacteria, archaea
ABW11964.1
Frankia sp EAN1pec
PspA PspAA TM(s) Metallo peptidase TM(s) PspAB
actinobacteria
AKB54760.1
Methanosarcina barkeri MS
PspA PspAA PspAB TM(s) Metallo peptidase TM(s)
archaea, chloroflexi, firmicutes, actinobacteria, proteobacteria
SIG SHOCT-bihelical TM(s) Metallo peptidase TM(s) PspAB PspA PspAA
AKX93460.1
Moorella thermoacetica
firmicutes
With ClgR, PspMN or Thioredoxin
AKK09009.1
Corynebacterium testudinoris
clgR-HTH PspA PspM PspN_N DUF3046
CCP45543.1
Mycobacterium tuberculosis H37Rv
actinobacteria
AKK09942.1
Corynebacterium testudinoris
Thioredoxin PspA
actinobacteria
With chaperones
ANH61663.1
Dokdonia donghaensis DSW1
CesT_Tir-DUF2170 PspA SIG TM(s) SIG Band-7 Coiled-coil Flot Betapropeller AAA-ATPase PhospholipaseD
FCB, proteobacteria, firmicutes
AAN56746.1
Shewanella oneidensis MR1
RHH DUF4178 PspA CesT_Tir-DUF2170 Spermine_synth TM(s) SIG YjfL-TM(s) Low-comp Ctha_1186 LipoSIG
γ-proteobacteria, bacteroidetes
CCG46123.1
Halobacillus halophilus DSM 2266
TM DUF4178 PspA SIG DUF4247 YjfL-TM(s) SIG TM(s) Spermine_synth
firmicutes
AAS43862.1
Bacillus cereus ATCC 10987
PspA YjfL-TM(s) SIG DUF4178 SIG DUF4247
firmicutes
BAB38581.1
Escherichia coli O157:H7 str Sakai
CesT_Tir-DUF2170 PspA SIG lon_trans_2 DUF2491 YjfL-TM(s) SIG DUF1190 SpermGS-ATPgrasp
γ-proteobacteria
ANL88610.1
Rhizobium phaseoli
PspA CesT_Tir-DUF2170
AEF02306.1
Alteromonas naphthalenivorans
α-proteobacteria
AAK24333.1
Caulobacter vibrioides CB15
CesT_Tir-DUF2170 PspA SpermGS-ATPgrasp
α-proteobacteria
ANX09535.1
Bacillus subtilis subsp subtilis
PspA ZnR(s) TM SIG TPM_phosphatase TM Band-7 ZnR
firmicutes
AKL67689.1
Streptomyces sp. Mg1
SIG TPM_phosphatase TM Coiled-coil PspA NYN REC wHTH SIG TM HAMP HISKIN
ALD73647.1
Trueperella pyogenes
actinobacteria, firmicutes, thermotogae
With two-component systems
ANX06812.1
Bacillus subtilis subsp subtilis
SIG Lial-LiaF-TM PspA SIG Toastrack Lial-LiaF-TM Toastrack SIG TM HAMP HISKIN REC wHTH
firmicutes
APB74393.1
Paenibacillus polymyxa
Lial-LiaF-TM PspA PspC Coiled-coil PspA SIG TM Toastrack SIG TM(s) HISKIN REC wHTH
firmicutes
Cyanobacterial and related contexts
AFZ14666.1
Crinalium epipsammum PCC 9333
PspA PspA Thioredoxin
cyanobacteria
BAG04642.1
Microcystis aeruginosa NIES843
PspA SIG TPR PBPI
cyanobacteria
AIK12556.1
Pectobacterium atrosepticum
SIG TPR TM DUF5401
γ-proteobacteria

Figure 4. Lineage spread of PspA-free domain architectures.

The domain architectures of the most prevalent homologs of PspA partner domains (frequency of occurrence >50 across lineages), including classical (Toastrack, LiaI-LiaF-TM, PspBC, PspMN, DUF3046) and other novel neighbors (PspAA, PspAB, HAAS, SHOCT-bihelical, SHOCT-like, AAA+-ATPase domains) are illustrated on the left. Broad distributions of domains across the tree of life are indicated in **Fig. 1C**. The phyletic spread of the underlying domain architectures is shown here along with their relative frequencies as a stacked barplot for all superkingdoms/domains of life (sub-lineages or phyla are sorted; *e.g.,* all bacterial lineages appear together). Further details of the domain architectures of all PspA partner domain homologs and their phyletic spreads are shown in the web app, with representatives shown in **Table S4**.

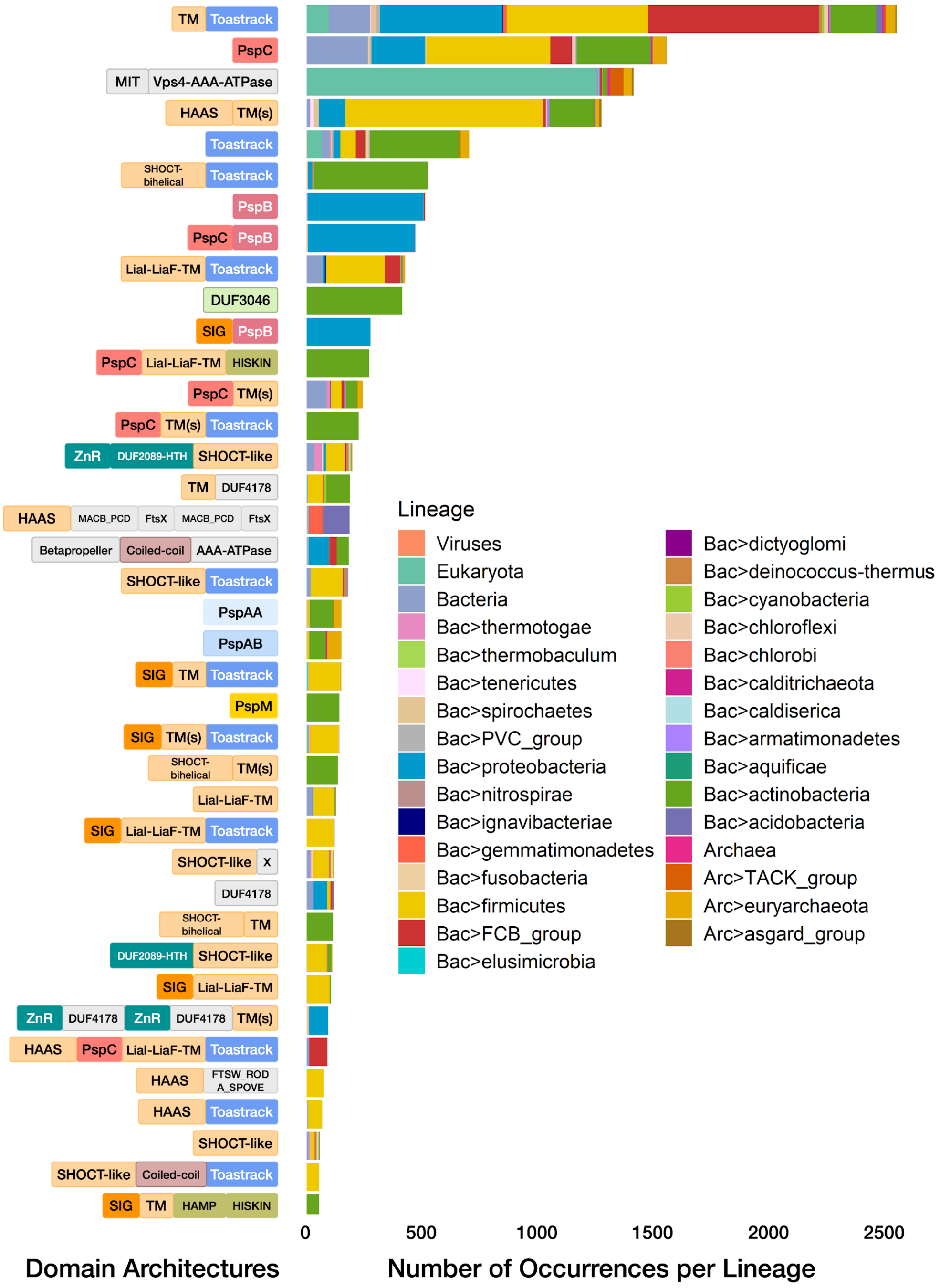

TM Toastrack
PspC
MIT Vps4-AAA-ATPase
HAAS TM(s)
Toastrack
SHOCT-bihelical Toastrack
PspB
PspC PspB
LiaI-LiaF-TM Toastrack
DUF3046
SIG PspB
PspC LiaI-LiaF-TM HISKIN
PspC TM(s)
PspC TM(s) Toastrack
ZnR DUF2089-HTH SHOCT-like
TM DUF4178
HAAS MACB_PCD FtsX MACB_PCD FtsX
Betapropeller Coiled-coil AAA-ATPase
SHOCT-like Toastrack
PspAA
PspAB
SIG TM Toastrack
PspM
SIG TM(s) Toastrack
SHOCT-bihelical TM(s)
LiaI-LiaF-TM
SIG LiaI-LiaF-TM Toastrack
SHOCT-like X
DUF4178
SHOCT-bihelical TM
DUF2089-HTH SHOCT-like
SIG LiaI-LiaF-TM
ZnR DUF4178 ZnR DUF4178 TM(s)
HAAS PspC LiaI-LiaF-TM Toastrack
HAAS FTSW_ROD A_SPOVE
HAAS Toastrack
SHOCT-like
SHOCT-like Coiled-coil Toastrack
SIG TM HAMP HISKIN
0
500
1000
1500
2000
2500
Domain Architectures
Number of Occurrences per Lineage
Lineage
Viruses
Eukaryota
Bacteria
Bac>thermotogae
Bac>thermobaculum
Bac>tenericutes
Bac>spirochaetes
Bac>PVC_group
Bac>proteobacteria
Bac>nitrospirae
Bac>ignavibacteriae
Bac>gemmatimonadetes
Bac>fusobacteria
Bac>firmicutes
Bac>FCB_group
Bac>elusimicrobia
Bac>dictyoglomi
Bac>deinococcus-thermus
Bac>cyanobacteria
Bac>chloroflexi
Bac>chlorobi
Bac>calditrichaeota
Bac>caldiserica
Bac>armatimonadetes
Bac>aquificae
Bac>actinobacteria
Bac>acidobacteria
Archaea
Arc>TACK_group
Arc>euryarchaeota
Arc>asgard_group

Figure 5. The genomic contexts housing all PSP cognate partner domain homologs.

The genomic contexts and key lineage memberships are presented using the same schematic as in **Figure 3**. The focus is on PSP partner domains such as Toastrack (blue), PspC, LiaI-LiaF-TM, HAAS, SHOCT-bi-helical (in shades of orange since they are transmembrane domains), and the various genomic neighborhoods with SHOCT-like proteins, transcription regulators (*e.g.,* PadR-wHTH, SIGMA-HTH, GNTR-HTH), and two-component systems (**Table S2**). Further details of the genomic contexts of all PspA-free partner domain homologs and their phyletic spreads are in the web app, and representatives indicated in the figure are shown in **Table S4**.

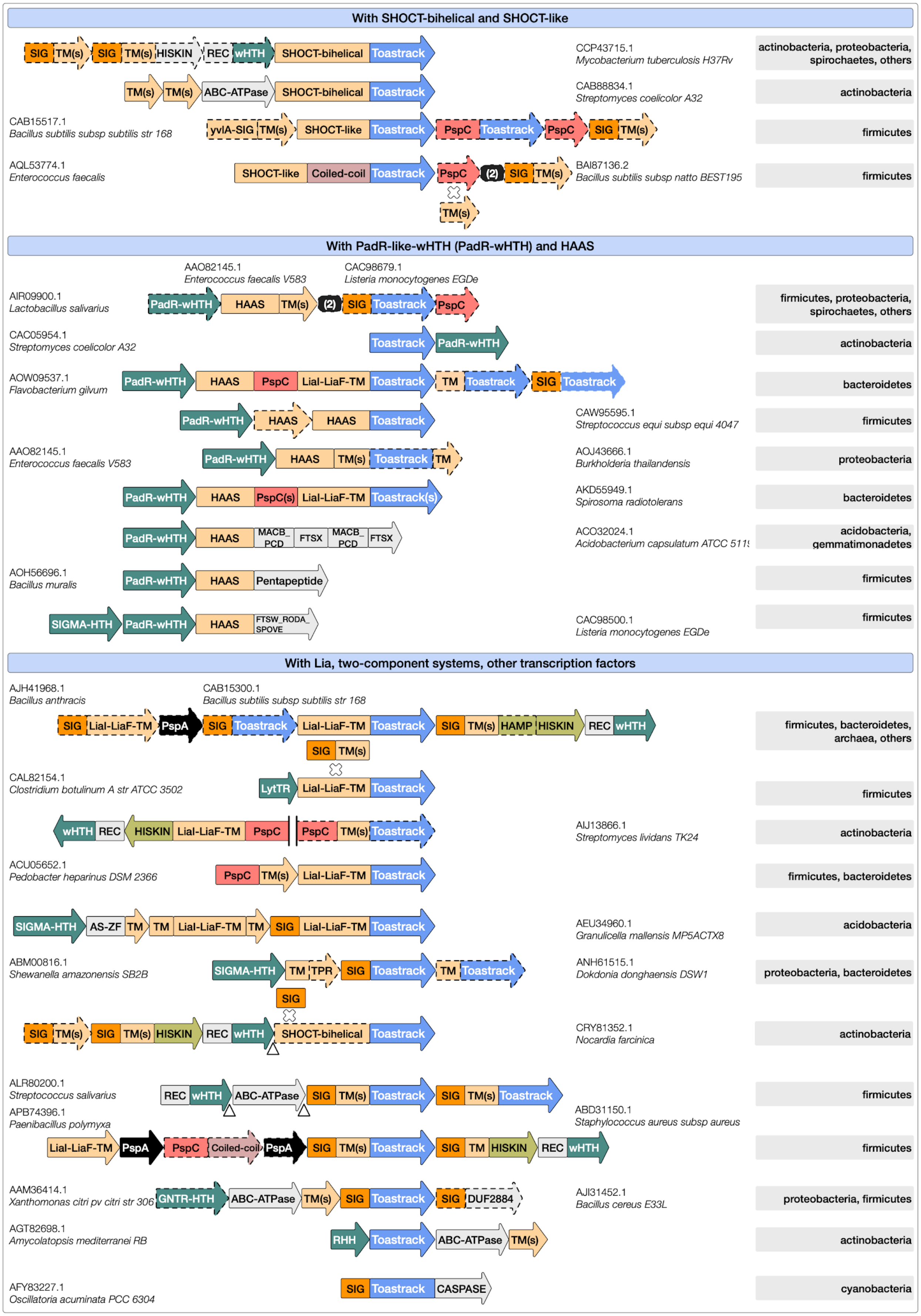
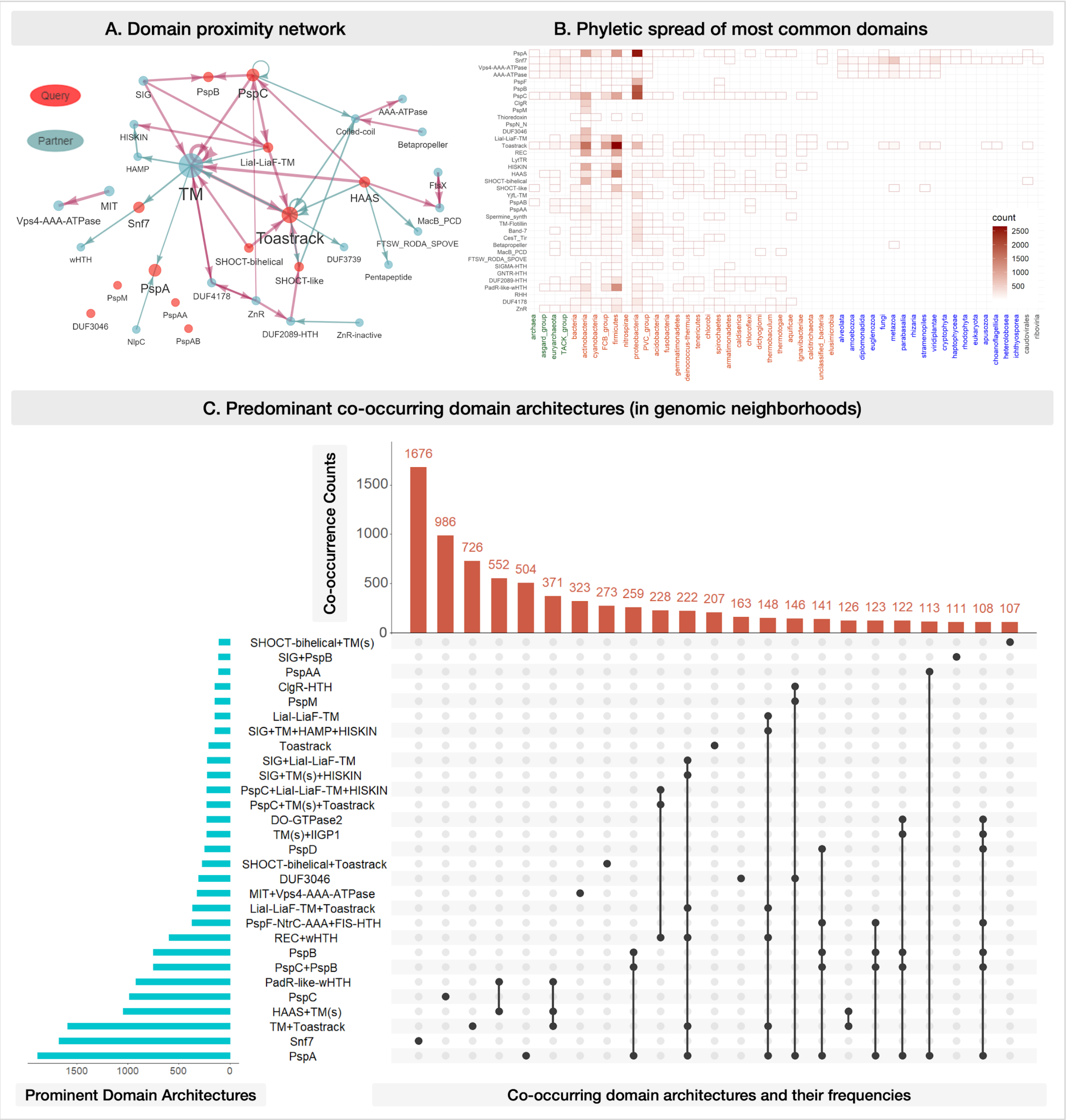

With SHOCT-bihelical and SHOCT-like
SIG TM(s) SIG TM(s) HISKIN REC wHTH SHOCT-bihelical Toastrack
CCP43715.1
Mycobacterium tuberculosis H37Rv
actinobacteria, proteobacteria, spirochaetes, others
TM(s) TM(s) ABC-ATPase SHOCT-bihelical Toastrack
CAB88834.1
Streptomyces coelicolor A32
actinobacteria
CAB15517.1
Bacillus subtilis subsp subtilis str 168
yvlA-SIG TM(s) SHOCT-like Toastrack PspC Toastrack PspC SIG TM(s)
firmicutes
AQL53774.1
Enterococcus faecalis
SHOCT-like Coiled-coil Toastrack PspC (2) SIG TM(s)
TM(s)
BAI87136.2
Bacillus subtilis subsp natto BEST195
firmicutes
With PadR-like-wHTH (PadR-wHTH) and HAAS
AAO82145.1
Enterococcus faecalis V583
CAC98679.1
Listeria monocytogenes EGDe
AIR09900.1
Lactobacillus salivarius
PadR-wHTH HAAS TM(s) (2) SIG Toastrack PspC
firmicutes, proteobacteria, spirochaetes, others
CAC05954.1
Streptomyces coelicolor A32
Toastrack PadR-wHTH
actinobacteria
AOW09537.1
Flavobacterium gilvum
PadR-wHTH HAAS PspC LiaI-LiaF-TM Toastrack TM Toastrack SIG Toastrack
bacteroidetes
PadR-wHTH HAAS HAAS Toastrack
CAW95595.1
Streptococcus equi subsp equi 4047
firmicutes
AAO82145.1
Enterococcus faecalis V583
PadR-wHTH HAAS TM(s) Toastrack TM
AOJ43666.1
Burkholderia thailandensis
proteobacteria
PadR-wHTH HAAS PspC(s) LiaI-LiaF-TM Toastrack(s)
AKD55949.1
Spirosoma radiotolerans
bacteroidetes
PadR-wHTH HAAS MACB_PCD FTSX MACB_PCD FTSX
ACO32024.1
Acidobacterium capsulatum ATCC 511
acidobacteria, gemmatimonadetes
AOH56696.1
Bacillus muralis
PadR-wHTH HAAS Pentapeptide
firmicutes
SIGMA-HTH PadR-wHTH HAAS FTSW_RODA_SPOVE
CAC98500.1
Listeria monocytogenes EGDe
firmicutes
With Lia, two-component systems, other transcription factors
AJH41968.1
Bacillus anthracis
CAB15300.1
Bacillus subtilis subsp subtilis str 168
SIG LiaI-LiaF-TM PspA SIG Toastrack LiaI-LiaF-TM Toastrack SIG TM(s) HAMP HISKIN REC wHTH
SIG TM(s)
firmicutes, bacteroidetes, archaea, others
CAL82154.1
Clostridium botulinum A str ATCC 3502
LytTR LiaI-LiaF-TM Toastrack
firmicutes
wHTH REC HISKIN LiaI-LiaF-TM PspC PspC TM(s) Toastrack
AIJ13866.1
Streptomyces lividans TK24
actinobacteria
ACU05652.1
Pedobacter heparinus DSM 2366
PspC TM(s) LiaI-LiaF-TM Toastrack
firmicutes, bacteroidetes
SIGMA-HTH AS-ZF TM TM LiaI-LiaF-TM TM SIG LiaI-LiaF-TM Toastrack
AEU34960.1
Granulicella mallensis MP5ACTX8
acidobacteria
ABM00816.1
Shewanella amazonensis SB2B
SIGMA-HTH TM TPR SIG Toastrack TM Toastrack
SIG
ANH61515.1
Dokdonia donghaensis DSW1
proteobacteria, bacteroidetes
SIG TM(s) SIG TM(s) HISKIN REC wHTH SHOCT-bihelical Toastrack
CRY81352.1
Nocardia farcinica
actinobacteria
ALR80200.1
Streptococcus salivarius
REC wHTH ABC-ATPase SIG TM(s) Toastrack SIG TM(s) Toastrack
firmicutes
APB74396.1
Paenibacillus polymyxa
LiaI-LiaF-TM PspA PspC Coiled-coil PspA SIG TM(s) Toastrack SIG TM HISKIN REC wHTH
ABD31150.1
Staphylococcus aureus subsp aureus
firmicutes
AAM36414.1
Xanthomonas citri pv citri str 306
GNTR-HTH ABC-ATPase TM(s) SIG Toastrack SIG DUF2884
AJI31452.1
Bacillus cereus E33L
proteobacteria, firmicutes
AGT82698.1
Amycolatopsis mediterranei RB
RHH Toastrack ABC-ATPase TM(s)
actinobacteria
AFY83227.1
Oscillatoria acuminata PCC 6304
SIG Toastrack CASPASE
cyanobacteria

Figure 6. PSP consolidated.

A. Domain proximity network. The network captures co-occurring domains within the top 97% of the homologs of all the 'query' Psp members and their key partner domains (after sorting by decreasing frequency of occurrence). The size of the nodes (domains) and width of edges (co-occurrence of domains within a protein) are proportional to the frequency of their occurrence across homologs. The query domains (original proteins/domains of interest; **Table S1**) and other commonly co-occurring domains (**Table S2**) are ***red*** or ***grey***. *Note*: A few connections may be absent in the displayed network due to low occurrence (fewer than the threshold), *e.g.,* PspA and PspAA, Betapropeller, and AAA-ATPase. The full network, and the domain-centric ones, can be accessed on the web app. **B. Phyletic spread of the most common domains.** The heatmap shows the presence/absence of homologs of PspA and partner domains across lineages. The color gradient represents the number of homologs identified within each lineage (*e.g.,* the darkest shade indicates the highest number of homologs in a particular lineage, white indicates the lowest count, and the absence of a shaded box represents a lack of available data). *Rows*: PSP members and their most frequent partners are queried against all sequenced and completed genomes across Bacteria, Eukaryota, and Archaea. *Columns*: The major archaeal (***green***), bacterial (***orange***), eukaryotic (***blue***), and viral (***grey***) lineages with representative sequenced genomes are considered. Details of all homologs across the tree of life, their domain architectures, genomic contexts, and their lineage distributions are shown in the web app (representatives in **Table S3, S4**). **C. Predominant co-occurring domain architectures in genomic neighborhoods.** UpSet plot of the most common neighboring proteins (genomic contexts >100 occurrences are shown) among all Psp homologs. *Blue histogram*: Distribution of the predominant domain architectures. *Dots and connections*: Combinations in which these domain architectures come together in the genomic neighborhoods. *Red histogram*: Frequency of occurrences of genomic neighborhoods comprising specific combinations of the predominant domain architectures. Phyletic spreads and UpSet plots of the domain architectures and genomic contexts for the homologs of all PSP member proteins are available in the web app.



Declarations

Acknowledgments

We are incredibly grateful to Arjun Krishnan and Krishnan Raghunathan for detailed comments on the manuscript and web application, guidance, and support; Emily Meyer and Jerome McKay for critical edits of the manuscript. We also thank Srinand Sreevatsan, Beronda Montgomery, Deborah Johnson, and the MSU Diversity Research Network for support and writing spaces. Finally, we thank members of the JRaviLab for feedback on the web application, and especially Jacob Krol and Faisal Alquaddoomi for their additional help with the final deployment. This work utilized the Alpine high-performance computing resource at the University of Colorado Boulder. Alpine is jointly funded by the University of Colorado Boulder, the University of Colorado Anschutz, Colorado State University, and the National Science Foundation (award: 2201538).

Funding

The following sources supported this work: National Institutes of Health (NIH) grants (R01AI104615, R01HL106788, and R01HL149450) awarded to M.L.G.; NIH Intramural Research Program awarded to V.A. and L.A; Oak Ridge Institute for Science and Education scholarship for the Visiting Scientist program to NIH, the Michigan State University College of Veterinary Medicine Endowed Research Funds, and University of Colorado start-up funds awarded to J.R.

Author Contributions

J.R. and M.L.G. conceived the study; J.R., V.A., L.A., and M.L.G. designed the study; J.R. and V.A acquired the data, performed all the analyses, and made the figures and tables. J.R., V.A., L.A., and M.L.G. interpreted the results and wrote the manuscript. S.Z.C built the web application with all the

results (data summarization and visualization) with J.R.; S.Z.C. also contributed to making figures and linking identifiers in the manuscript to reference databases. E.P.B. contributed to writing the manuscript, performed phylogenetic tree construction, and mapped Pfam to Interpro domain IDs throughout the manuscript. P.D. contributed to renaming some domains.

Data Availability and Reuse

All the data, analyses, and visualizations are available in our interactive and queryable web application: https://jravilab.org/psp. Text, figures, and the web app are licensed under Creative Commons Attribution CC BY 4.0.

References

1. van den Brink-van der Laan E, Killian JA, de Kruijff B. 2004. Nonbilayer lipids affect peripheral and integral membrane proteins via changes in the lateral pressure profile. Biochim Biophys Acta 1666:275–288.
2. Kobayashi R, Suzuki T, Yoshida M. 2007. Escherichia coli phage-shock protein A (PspA) binds to membrane phospholipids and repairs proton leakage of the damaged membranes. Mol Microbiol 66:100–109.
3. Jordan S, Hutchings MI, Mascher T. 2008. Cell envelope stress response in Gram-positive bacteria. FEMS Microbiol Rev 32:107–146.
4. Raivio TL. 2005. Envelope stress responses and Gram-negative bacterial pathogenesis. Mol Microbiol 56:1119–1128.
5. Rowley G, Spector M, Kormanec J, Roberts M. 2006. Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens. Nat Rev Microbiol 4:383–394.
6. Omardien S, Drijfhout JW, van Veen H, Schachtschabel S, Riool M, Hamoen LW, Brul S, Zaat SAJ. 2018. Synthetic antimicrobial peptides delocalize membrane bound proteins thereby inducing a cell envelope stress response. Biochim Biophys Acta Biomembr 1860:2416–2427.
7. Wenzel M, Kohl B, Münch D, Raatschen N, Albada HB, Hamoen L, Metzler-Nolte N, Sahl H-G, Bandow JE. 2012. Proteomic response of Bacillus subtilis to lantibiotics reflects differences in interaction with the cytoplasmic membrane. Antimicrob Agents Chemother 56:5749–5757.
8. Guest RL, Wang J, Wong JL, Raivio TL. 2017. A Bacterial Stress Response Regulates Respiratory Protein Complexes To Control Envelope Stress Adaptation. J Bacteriol 199.
9. Gottesman S. 2017. Stress Reduction, Bacterial Style. J Bacteriol 199.
10. Brissette JL, Weiner L, Ripmaster TL, Model P. 1991. Characterization and sequence of the Escherichia coli stress-induced psp operon. J Mol Biol 220:35–48.
11. Mascher T, Zimmer SL, Smith T-A, Helmann JD. 2004. Antibiotic-inducible promoter regulated by the cell envelope stress-sensing two-component system LiaRS of Bacillus subtilis. Antimicrob Agents Chemother 48:2888–2896.
12. Huvet M, Toni T, Sheng X, Thorne T, Jovanovic G, Engl C, Buck M, Pinney JW, Stumpf MPH. 2011. The evolution of the phage shock protein response system: interplay between protein function, genomic organization, and system function. Mol Biol Evol 28:1141–1155.
13. McDonald C, Jovanovic G, Ces O, Buck M. 2015. Membrane Stored Curvature Elastic Stress Modulates Recruitment of Maintenance Proteins PspA and Vipp1. mBio 6:e01188-01115.
14. Datta P, Ravi J, Guerrini V, Chauhan R, Neiditch MB, Shell SS, Fortune SM, Hancioglu B, Igoshin OA, Gennaro ML. 2015. The Psp system of Mycobacterium tuberculosis integrates envelope stress-sensing and envelope-preserving functions. Mol Microbiol 97:408–422.
15. Ravi J, Anantharaman V, Aravind L, Gennaro ML. 2018. Variations on a theme: evolution of the phage-shock-protein system in Actinobacteria. Antonie Van Leeuwenhoek 111:753–760.
16. Heidrich J, Thurotte A, Schneider D. 2017. Specific interaction of IM30/Vipp1 with cyanobacterial and chloroplast membranes results in membrane remodeling and eventually in membrane fusion. Biochim Biophys Acta Biomembr 1859:537–549.
17. McDonald C, Jovanovic G, Wallace BA, Ces O, Buck M. 2017. Structure and function of PspA and Vipp1 N-terminal peptides: Insights into the membrane stress sensing and mitigation. Biochim Biophys Acta Biomembr 1859:28–39.

18. Lee I-H, Kai H, Carlson L-A, Groves JT, Hurley JH. 2015. Negative membrane curvature catalyzes nucleation of endosomal sorting complex required for transport (ESCRT)-III assembly. Proc Natl Acad Sci U S A 112:15892–15897.
19. Liu J, Tassinari M, Souza DP, Naskar S, Noel JK, Bohuszewicz O, Buck M, Williams TA, Baum B, Low HH. 2021. Bacterial Vipp1 and PspA are members of the ancient ESCRT-III membrane-remodeling superfamily. Cell 184:3660-3673.e18.
20. Junglas B, Huber ST, Heidler T, Schlösser L, Mann D, Hennig R, Clarke M, Hellmann N, Schneider D, Sachse C. 2021. PspA adopts an ESCRT-III-like fold and remodels bacterial membranes. Cell 184:3674-3688.e18.
21. Horinouchi T, Suzuki S, Hirasawa T, Ono N, Yomo T, Shimizu H, Furusawa C. 2015. Phenotypic convergence in bacterial adaptive evolution to ethanol stress. BMC Evol Biol 15:180.
22. Furusawa C, Irie N. 2020. Toward understanding of evolutionary constraints: experimental and theoretical approaches. Biophys Rev 12:1155–1161.
23. Kleerebezem M, Crielaard W, Tommassen J. 1996. Involvement of stress protein PspA (phage shock protein A) of Escherichia coli in maintenance of the protonmotive force under stress conditions. EMBO J 15:162–171.
24. Huvet M, Toni T, Tan H, Jovanovic G, Engl C, Buck M, Stumpf MPH. 2009. Model-based evolutionary analysis: the natural history of phage-shock stress response. Biochem Soc Trans 37:762–767.
25. Vothknecht UC, Otters S, Hennig R, Schneider D. 2012. Vipp1: a very important protein in plastids?! J Exp Bot 63:1699–1712.
26. Bultema JB, Fuhrmann E, Boekema EJ, Schneider D. 2010. Vipp1 and PspA: Related but not twins. Commun Integr Biol 3:162–165.
27. Westphal S, Heins L, Soll J, Vothknecht UC. 2001. Vipp1 deletion mutant of Synechocystis: a connection between bacterial phage shock and thylakoid biogenesis? Proc Natl Acad Sci U S A 98:4243–4248.
28. Heidrich J, Wulf V, Hennig R, Saur M, Markl J, Sönnichsen C, Schneider D. 2016. Organization into Higher Ordered Ring Structures Counteracts Membrane Binding of IM30, a Protein Associated with Inner Membranes in Chloroplasts and Cyanobacteria. J Biol Chem 291:14954–14962.
29. Theis J, Gupta TK, Klingler J, Wan W, Albert S, Keller S, Engel BD, Schroda M. 2019. VIPP1 rods engulf membranes containing phosphatidylinositol phosphates. 1. Sci Rep 9:8725.
30. Zhang L, Kato Y, Otters S, Vothknecht UC, Sakamoto W. 2012. Essential role of VIPP1 in chloroplast envelope maintenance in Arabidopsis. Plant Cell 24:3695–3707.
31. Armstrong RM, Adams KL, Zilisch JE, Bretl DJ, Sato H, Anderson DM, Zahrt TC. 2016. Rv2744c Is a PspA Ortholog That Regulates Lipid Droplet Homeostasis and Nonreplicating Persistence in Mycobacterium tuberculosis. J Bacteriol 198:1645–1661.
32. Gupta TK, Klumpe S, Gries K, Heinz S, Wietrzynski W, Ohnishi N, Niemeyer J, Spaniol B, Schaffer M, Rast A, Ostermeier M, Strauss M, Plitzko JM, Baumeister W, Rudack T, Sakamoto W, Nickelsen J, Schuller JM, Schroda M, Engel BD. 2021. Structural basis for VIPP1 oligomerization and maintenance of thylakoid membrane integrity. Cell 184:3643-3659.e23.
33. Joly N, Engl C, Jovanovic G, Huvet M, Toni T, Sheng X, Stumpf MPH, Buck M. 2010. Managing membrane stress: the phage shock protein (Psp) response, from molecular mechanisms to physiology. FEMS Microbiol Rev 34:797–827.

34. Eiamphungporn W, Helmann JD. 2008. The Bacillus subtilis sigma(M) regulon and its contribution to cell envelope stress responses. Mol Microbiol 67:830–848.

35. Jordan S, Junker A, Helmann JD, Mascher T. 2006. Regulation of LiaRS-dependent gene expression in bacillus subtilis: identification of inhibitor proteins, regulator binding sites, and target genes of a conserved cell envelope stress-sensing two-component system. J Bacteriol 188:5153–5166.

36. Manganelli R, Gennaro ML. 2017. Protecting from Envelope Stress: Variations on the Phage-Shock-Protein Theme. Trends Microbiol 25:205–216.

37. Popp PF, Gumerov VM, Andrianova EP, Bewersdorf L, Mascher T, Zhulin IB, Wolf D. 2022. Phyletic Distribution and Diversification of the Phage Shock Protein Stress Response System in Bacteria and Archaea. mSystems 7:e01348-21.

38. van der Laan M, Urbanus ML, Ten Hagen-Jongman CM, Nouwen N, Oudega B, Harms N, Driessen AJM, Luirink J. 2003. A conserved function of YidC in the biogenesis of respiratory chain complexes. Proc Natl Acad Sci U S A 100:5801–5806.

39. Maxson ME, Darwin AJ. 2006. Multiple promoters control expression of the Yersinia enterocolitica phage-shock-protein A (pspA) operon. Microbiol Read Engl 152:1001–1010.

40. Weiner L, Model P. 1994. Role of an Escherichia coli stress-response operon in stationary-phase survival. Proc Natl Acad Sci U S A 91:2191–2195.

41. Jovanovic G, Lloyd LJ, Stumpf MPH, Mayhew AJ, Buck M. 2006. Induction and function of the phage shock protein extracytoplasmic stress response in Escherichia coli. J Biol Chem 281:21147–21161.

42. Mitchell AM, Silhavy TJ. 2019. Envelope stress responses: balancing damage repair and toxicity. Nat Rev Microbiol 17:417–428.

43. Engl C, Beek AT, Bekker M, de Mattos JT, Jovanovic G, Buck M. 2011. Dissipation of proton motive force is not sufficient to induce the phage shock protein response in Escherichia coli. Curr Microbiol 62:1374–1385.

44. Weiner L, Brissette JL, Model P. 1991. Stress-induced expression of the Escherichia coli phage shock protein operon is dependent on sigma 54 and modulated by positive and negative feedback mechanisms. Genes Dev 5:1912–1923.

45. Jovanovic G, Engl C, Buck M. 2009. Physical, functional and conditional interactions between ArcAB and phage shock proteins upon secretin-induced stress in Escherichia coli. Mol Microbiol 74:16–28.

46. Yamaguchi S, Reid DA, Rothenberg E, Darwin AJ. 2013. Changes in Psp protein binding partners, localization and behaviour upon activation of the Yersinia enterocolitica phage shock protein response. Mol Microbiol 87:656–671.

47. Wolf D, Kalamorz F, Wecke T, Juszczak A, Mäder U, Homuth G, Jordan S, Kirstein J, Hoppert M, Voigt B, Hecker M, Mascher T. 2010. In-depth profiling of the LiaR response of Bacillus subtilis. J Bacteriol 192:4680–4693.

48. Jordan S, Rietkötter E, Strauch MA, Kalamorz F, Butcher BG, Helmann JD, Mascher T. 2007. LiaRS-dependent gene expression is embedded in transition state regulation in Bacillus subtilis. Microbiol Read Engl 153:2530–2540.

49. Castelle CJ, Banfield JF. 2018. Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. Cell 172:1181–1197.

50. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nat Microbiol 1:16048.

51. Flores-Kim J, Darwin AJ. 2016. The Phage Shock Protein Response. Annu Rev Microbiol 70:83–101.

52. Brissette JL, Russel M, Weiner L, Model P. 1990. Phage shock protein, a stress protein of Escherichia coli. Proc Natl Acad Sci U S A 87:862–866.

53. Darwin AJ, Miller VL. 2001. The psp locus of Yersinia enterocolitica is required for virulence and for growth in vitro when the Ysc type III secretion system is produced. Mol Microbiol 39:429–444.

54. Becker LA, Bang I-S, Crouch M-L, Fang FC. 2005. Compensatory role of PspA, a member of the phage shock protein operon, in rpoE mutant Salmonella enterica serovar Typhimurium. Mol Microbiol 56:1004–1016.

55. Elderkin S, Bordes P, Jones S, Rappas M, Buck M. 2005. Molecular determinants for PspA-mediated repression of the AAA transcriptional activator PspF. J Bacteriol 187:3238–3248.

56. Suntharalingam P, Senadheera MD, Mair RW, Lévesque CM, Cvitkovitch DG. 2009. The LiaFSR system regulates the cell envelope stress response in Streptococcus mutans. J Bacteriol 191:2973–2984.

57. Aravind L, Ponting CP. 1999. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. FEMS Microbiol Lett 176:111–116.

58. Flores-Kim J, Darwin AJ. 2012. Phage shock protein C (PspC) of Yersinia enterocolitica is a polytopic membrane protein with implications for regulation of the Psp stress response. J Bacteriol 194:6548–6559.

59. Burki F, Shalchian-Tabrizi K, Pawlowski J. 2008. Phylogenomics reveals a new “megagroup” including most photosynthetic eukaryotes. Biol Lett 4:366–369.

60. Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AJ. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “supergroups.” Proc Natl Acad Sci U S A 106:3859–3864.

61. Caillat C, Maity S, Miguet N, Roos WH, Weissenhorn W. 2019. The role of VPS4 in ESCRT-III polymer remodeling. Biochem Soc Trans 47:441–448.

62. Maxson ME, Darwin AJ. 2006. PspB and PspC of Yersinia enterocolitica are dual function proteins: regulators and effectors of the phage-shock-protein response. Mol Microbiol 59:1610–1623.

63. Gueguen E, Savitzky DC, Darwin AJ. 2009. Analysis of the Yersinia enterocolitica PspBC proteins defines functional domains, essential amino acids and new roles within the phage-shock-protein response. Mol Microbiol 74:619–633.

64. Wallrodt I, Jelsbak L, Thomsen LE, Brix L, Lemire S, Gautier L, Nielsen DS, Jovanovic G, Buck M, Olsen JE. 2014. Removal of the phage-shock protein PspB causes reduction of virulence in Salmonella enterica serovar Typhimurium independently of NRAMP1. J Med Microbiol 63:788–795.

65. Kleine B, Chattopadhyay A, Polen T, Pinto D, Mascher T, Bott M, Brocker M, Freudl R. 2017. The three-component system EsrISR regulates a cell envelope stress response in Corynebacterium glutamicum. Mol Microbiol 106:719–741.

66. Srivastava D, Moumene A, Flores-Kim J, Darwin AJ. 2017. Psp Stress Response Proteins Form a Complex with Mislocalized Secretins in the Yersinia enterocolitica Cytoplasmic Membrane. mBio 8.

67. Datta P, Ravi J, Guerrini V, Chauhan R, Neiditch MB, Shell SS, Fortune SM, Hancioglu B, Igoshin OA, Gennaro ML. 2015. The Psp system of Mycobacterium tuberculosis integrates envelope stress-sensing and envelope-preserving functions. Mol Microbiol 97:408–422.

68. Schöneberg J, Pavlin MR, Yan S, Righini M, Lee I-H, Carlson L-A, Bahrami AH, Goldman DH, Ren X, Hummer G, Bustamante C, Hurley JH. 2018. ATP-dependent force generation and membrane scission by ESCRT-III and Vps4. Science 362:1423–1428.

69. Weiss P, Huppert S, Kölling R. 2009. Analysis of the dual function of the ESCRT-III protein Snf7 in endocytic trafficking and in gene expression. Biochem J 424:89–97.

70. Godinho RM da C, Crestani J, Kmetzsch L, Araujo G de S, Frases S, Staats CC, Schrank A, Vainstein MH, Rodrigues ML. 2014. The vacuolar-sorting protein Snf7 is required for export of virulence determinants in members of the Cryptococcus neoformans complex. Sci Rep 4:6198.

71. Webster BM, Colombi P, Jäger J, Lusk CP. 2014. Surveillance of nuclear pore complex assembly by ESCRT-III/Vps4. Cell 159:388–401.

72. Zhang L, Kondo H, Kamikubo H, Kataoka M, Sakamoto W. 2016. VIPP1 Has a Disordered C-Terminal Tail Necessary for Protecting Photosynthetic Membranes against Stress. Plant Physiol 171:1983–1995.

73. Anantharaman V, Aravind L. 2003. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome Biol 4:R11.

74. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST. 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A 99:3684–3689.

75. Tang S, Henne WM, Borbat PP, Buchkovich NJ, Freed JH, Mao Y, Fromme JC, Emr SD. 2015. Structural basis for activation, assembly and membrane binding of ESCRT-III Snf7 filaments. eLife 4.

76. Anantharaman V, Koonin EV, Aravind L. 2001. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol 307:1271–1292.

77. Kaur G, Burroughs AM, Iyer LM, Aravind L. 2020. Highly regulated, diversifying NTP-dependent biological conflict systems with implications for the emergence of multicellularity. eLife 9.

78. Tsai JT, Sung N, Lee J, Chang C, Lee S, Tsai FTF. 2019. Crystal Structure of the YcjX Stress Protein Reveals a Ras-Like GTP-Binding Protein. J Mol Biol 431:3179–3190.

79. Wollert T, Wunder C, Lippincott-Schwartz J, Hurley JH. 2009. Membrane scission by the ESCRT-III complex. Nature 458:172–177.

80. Frohn BP, Härtel T, Cox J, Schwille P. 2022. Tracing back variations in archaeal ESCRT-based cell division to protein domain architectures. PloS One 17:e0266395.

81. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU, Stott MB, Nunoura T, Banfield JF, Schramm A, Baker BJ, Spang A, Ettema TJG. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541:353–358.

82. Burroughs AM, Aravind L. 2016. RNA damage in biological conflicts and the diversity of responding RNA repair systems. Nucleic Acids Res 44:8525–8555.

83. Langhorst MF, Reuter A, Stuermer CAO. 2005. Scaffolding microdomains and beyond: the function of reggie/flotillin proteins. Cell Mol Life Sci 62:2228–2240.

84. Gehl B, Sweetlove LJ. 2014. Mitochondrial Band-7 family proteins: scaffolds for respiratory chain assembly? Front Plant Sci 5:141.

85. Delahay RM, Shaw RK, Elliott SJ, Kaper JB, Knutton S, Frankel G. 2002. Functional analysis of the enteropathogenic Escherichia coli type III secretion system chaperone CesT identifies domains that mediate substrate interactions. Mol Microbiol 43:61–73.

86. Jovanovic G, Mehta P, McDonald C, Davidson AC, Uzdavinys P, Ying L, Buck M. 2014. The N-terminal amphipathic helices determine regulatory and effector functions of phage shock protein A (PspA) in Escherichia coli. J Mol Biol 426:1498–1511.

87. Lee YH, Kingston AW, Helmann JD. 2012. Glutamate dehydrogenase affects resistance to cell wall antibiotics in Bacillus subtilis. J Bacteriol 194:993–1001.

88. Jovanovic G, Mehta P, Ying L, Buck M. 2014. Anionic lipids and the cytoskeletal proteins MreB and RodZ define the spatio-temporal distribution and function of membrane stress controller PspA in Escherichia coli. Microbiol Read Engl 160:2374–2386.

89. Lee M-J, Huang C-Y, Sun Y-J, Huang H. 2005. Cloning and characterization of spermidine synthase and its implication in polyamine biosynthesis in Helicobacter pylori strain 26695. Protein Expr Purif 43:140–148.

90. Li B, Kurihara S, Kim SH, Liang J, Michael AJ. 2019. A polyamine-independent role for S-adenosylmethionine decarboxylase. Biochem J 476:2579–2594.

91. Cerutti ML, Otero LH, Smal C, Pellizza L, Goldbaum FA, Klinke S, Aran M. 2017. Structural and functional characterization of a cold adapted TPM-domain with ATPase/ADPase activity. J Struct Biol 197:201–209.

92. Wegener KM, Bennewitz S, Oelmüller R, Pakrasi HB. 2011. The Psb32 protein aids in repairing photodamaged photosystem II in the cyanobacterium Synechocystis 6803. Mol Plant 4:1052–1061.

93. Sirpiö S, Allahverdiyeva Y, Suorsa M, Paakkarinen V, Vainonen J, Battchikova N, Aro E-M. 2007. TLP18.3, a novel thylakoid lumen protein regulating photosystem II repair cycle. Biochem J 406:415–425.

94. Ballas SK, Mohandas N, Marton LJ, Shohet SB. 1983. Stabilization of erythrocyte membranes by polyamines. Proc Natl Acad Sci U S A 80:1942–1946.

95. Fiuza M, Letek M, Leiba J, Villadangos AF, Vaquera J, Zanella-Cléon I, Mateos LM, Molle V, Gil JA. 2010. Phosphorylation of a novel cytoskeletal protein (RsmP) regulates rod-shaped morphology in Corynebacterium glutamicum. J Biol Chem 285:29387–29397.

96. Eberhardt RY, Bartholdson SJ, Punta M, Bateman A. 2013. The SHOCT domain: a widespread domain under-represented in model organisms. PloS One 8:e57848.

97. Kingston AW, Liao X, Helmann JD. 2013. Contributions of the σ(W) , σ(M) and σ(X) regulons to the lantibiotic resistome of Bacillus subtilis. Mol Microbiol 90:502–518.

98. Miller C, Kong J, Tran TT, Arias CA, Saxer G, Shamoo Y. 2013. Adaptation of Enterococcus faecalis to daptomycin reveals an ordered progression to resistance. Antimicrob Agents Chemother 57:5373–5383.

99. Wang L, Xiao S, Chen X, Chen S, Wang S, Wang C, Tang Y, Dong F. 2019. ytiB and ythA Genes Reduce the Uranium Removal Capacity of Bacillus atrophaeus. Int J Mol Sci 20.

100. Boyle-Vavra S, Yin S, Jo DS, Montgomery CP, Daum RS. 2013. VraT/YvqF is required for methicillin resistance and activation of the VraSR regulon in Staphylococcus aureus. Antimicrob Agents Chemother 57:83–95.

101. Gardete S, Wu SW, Gill S, Tomasz A. 2006. Role of VraSR in antibiotic resistance and antibiotic-induced stress response in Staphylococcus aureus. Antimicrob Agents Chemother 50:3424–3434.

102. Wolf D, Domínguez-Cuevas P, Daniel RA, Mascher T. 2012. Cell envelope stress response in cell wall-deficient L-forms of Bacillus subtilis. Antimicrob Agents Chemother 56:5907–5915.

103. Hauf S, Möller L, Fuchs S, Halbedel S. 2019. PadR-type repressors controlling production of a non-canonical FtsW/RodA homologue and other trans-membrane proteins. Sci Rep 9:10023.

104. Krol JD, Burke JT, Chen SZ, Sosinski L, Alquaddoomi FS, Brenner EP, Wolfe EP, Rubinetti VP, Amolitos S, Reason KM, Johnston JB, Ravi J. 2022. MolEvolvR: A web-app for characterizing proteins using molecular evolution and phylogeny. preprint. bioRxiv [Preprint] https://doi.org/10.1101/2022.02.18.461833.

105. Leipe DD, Wolf YI, Koonin EV, Aravind L. 2002. Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol 317:41–72.

106. Lane N, Martin WF. 2012. The origin of membrane bioenergetics. Cell 151:1406–1416.

107. Sojo V, Pomiankowski A, Lane N. 2014. A bioenergetic basis for membrane divergence in archaea and bacteria. PLoS Biol 12:e1001926.

108. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733-745.

109. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2016. GenBank. Nucleic Acids Res 44:D67–D72.

110. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, Dempsey DM, Dickerman A, Dietrich EM, Kenyon RW, Kuscuoglu M, Lefkowitz EJ, Lu J, Machi D, Macken C, Mao C, Niewiadomska A, Nguyen M, Olsen GJ, Overbeek JC, Parrello B, Parrello V, Porter JS, Pusch GD, Shukla M, Singh I, Stewart L, Tan G, Thomas C, VanOeffelen M, Vonstein V, Wallace ZS, Warren AS, Wattam AR, Xia F, Yoo H, Zhang Y, Zmasek CM, Scheuermann RH, Stevens RL. 2023. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. Nucleic Acids Res 51:D678–D689.

111. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res 45:D535–D542.

112. Federhen S. 2012. The NCBI Taxonomy database. Nucleic Acids Res 40:D136-143.

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

114. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244-248.

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

116. Nielsen H. 2017. Predicting Secretory Proteins with SignalP. Methods Mol Biol Clifton NJ 1611:59–73.

117. Käll L, Krogh A, Sonnhammer ELL. 2004. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338:1027–1036.

118. Käll L, Krogh A, Sonnhammer ELL. 2007. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res 35:W429-432.

119. Cole C, Barber JD, Barton GJ. 2008. The Jpred 3 secondary structure prediction server. Nucleic Acids Res 36:W197-201.

120. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432.

121. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR. 2015. HMMER web server: 2015 update. Nucleic Acids Res 43:W30-38.

122. Lassmann T. 2019. Kalign 3: multiple sequence alignment of large data sets. Bioinforma Oxf Engl https://doi.org/10.1093/bioinformatics/btz795.

123. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

124. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PloS One 5:e9490.

125. Rambaut A. 2018. FigTree v1.4.4. GitHub. Repository. https://github.com/rambaut/figtree/releases. Retrieved 13 July 2023.

126. Flouri T, Izquierdo-Carrasco F, Darriba D, Aberer AJ, Nguyen L-T, Minh BQ, Von Haeseler A, Stamatakis A. 2015. The Phylogenetic Likelihood Library. Syst Biol 64:356–362.

127. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Mol Biol Evol 37:291–294.

128. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455.

129. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. 2019. shiny: Web Application Framework for R. https://CRAN.R-project.org/package=shiny.

130. R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

131. RStudio Team. 2020. RStudio: Integrated Development Environment for R. RStudio, PBC., Boston, MA. http://www.rstudio.com/.

132. Hao ZhuThomas Travison DM Timothy Tsai, Will Beasley, Yihui Xie, GuangChuang Yu, Stéphane Laurent, Rob Shepherd, Yoni Sidi, Brian Salzer, George Gui, Yeliang Fan. 2020.

Construct Complex Table with “kable” and Pipe Syntax. http://haozhu233.github.io/kableExtra/, https://github.com/haozhu233/kableExtra.

133. Mike Bostock CY Kerry Rodden, Kevin Warne, Kent Russell, Florian Breitwieser. 2020. Sunburst “Htmlwidget.” https://github.com/timelyportfolio/sunburstR.

134. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. 2014. UpSet: Visualization of Intersecting Sets. IEEE Trans Vis Comput Graph 20:1983–1992.

135. Enrico Bonatesta UB Christoph Horejs-Kainrath. 2020. Multiple Sequence Alignment. http://www.bioinf.jku.at/software/msa/.

136. Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinforma Oxf Engl 35:526–528.

137. Fellows I. 2018. Word Clouds. http://blog.fellstat.com/?cat=11.

138. Jake Conway NG. 2019. A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets. http://github.com/hms-dbmi/UpSetR.

139. Müller K. 2017. here: A Simpler Way to Find Your Files. https://CRAN.R-project.org/package=here.

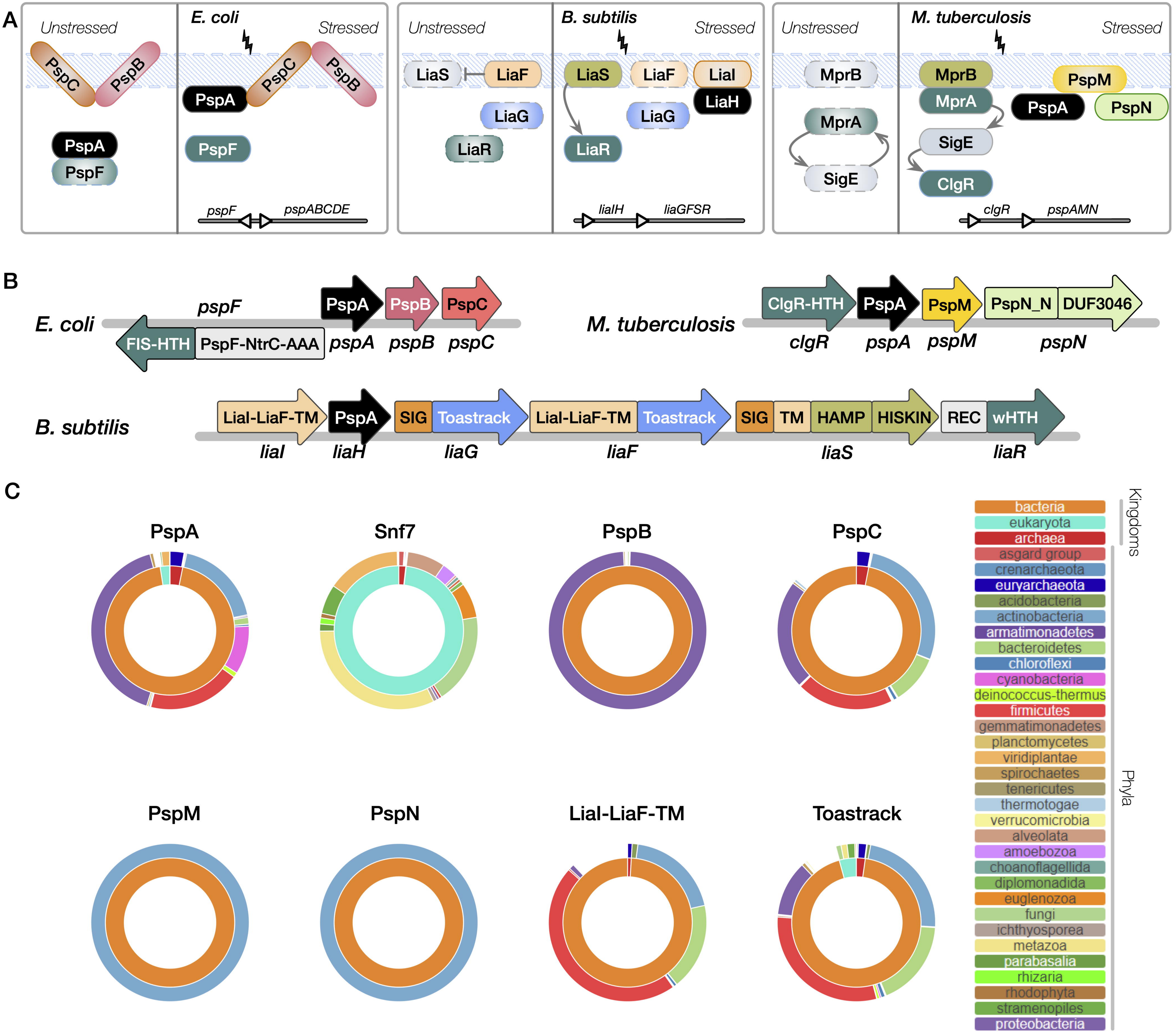
140. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the Tidyverse. J Open Source Softw 4:1686.

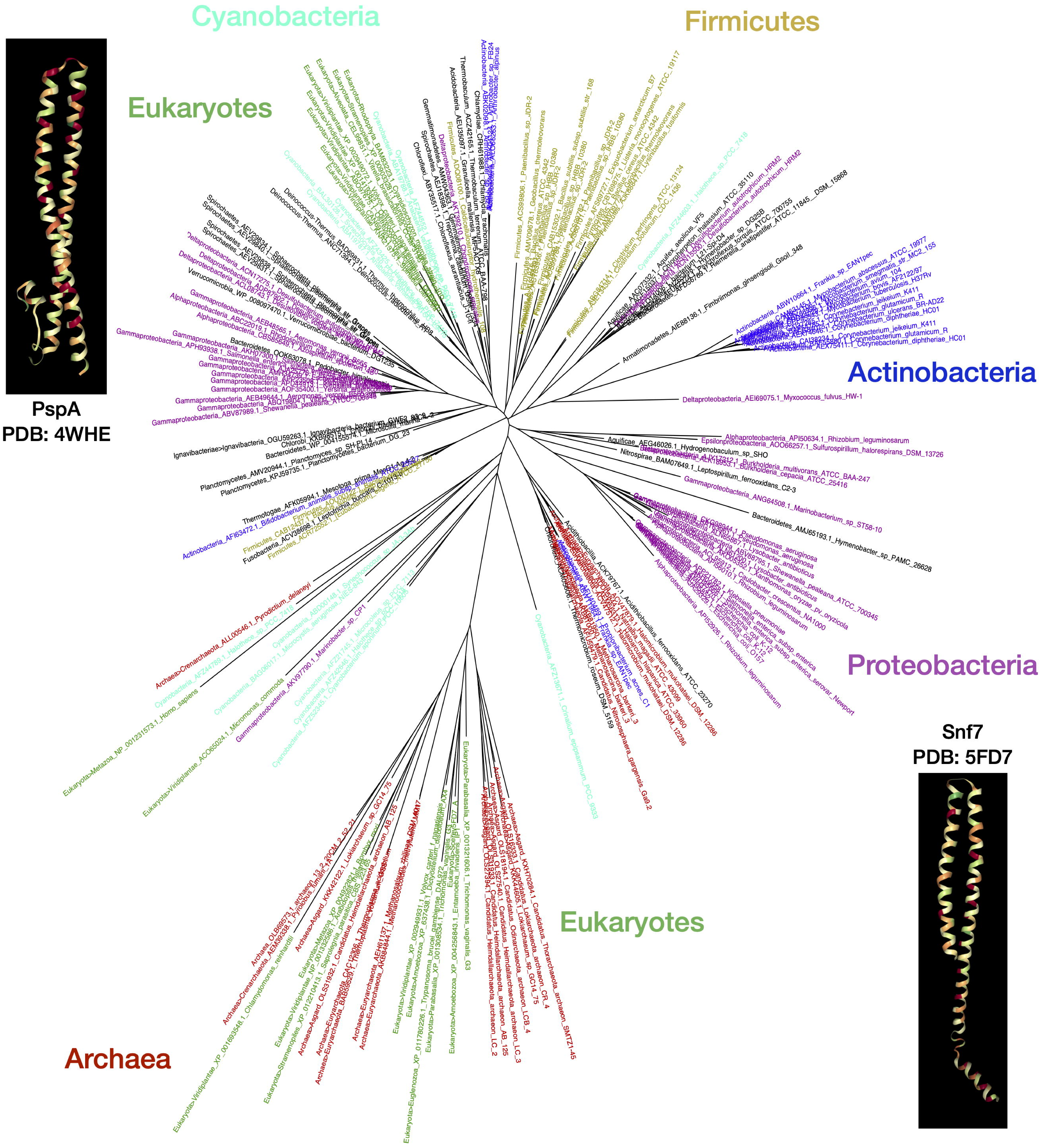
141. Martin Morgan MR. 2019. Access the Bioconductor Project Package Repository. https://github.com/Bioconductor/BiocManager.

142. Perry M. 2016. Flexible Heatmaps for Functional Genomics and Sequence Features. https://git.bioconductor.org/packages/heatmaps.

143. Xie Y, Allaire JJ, Grolemund G. 2018. R Markdown: The Definitive Guide. Chapman and Hall/CRC, Boca Raton, Florida. https://bookdown.org/yihui/rmarkdown.

144. Allaire JJ, Xie, Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R. 2020. rmarkdown: Dynamic Documents for R. https://github.com/rstudio/rmarkdown.

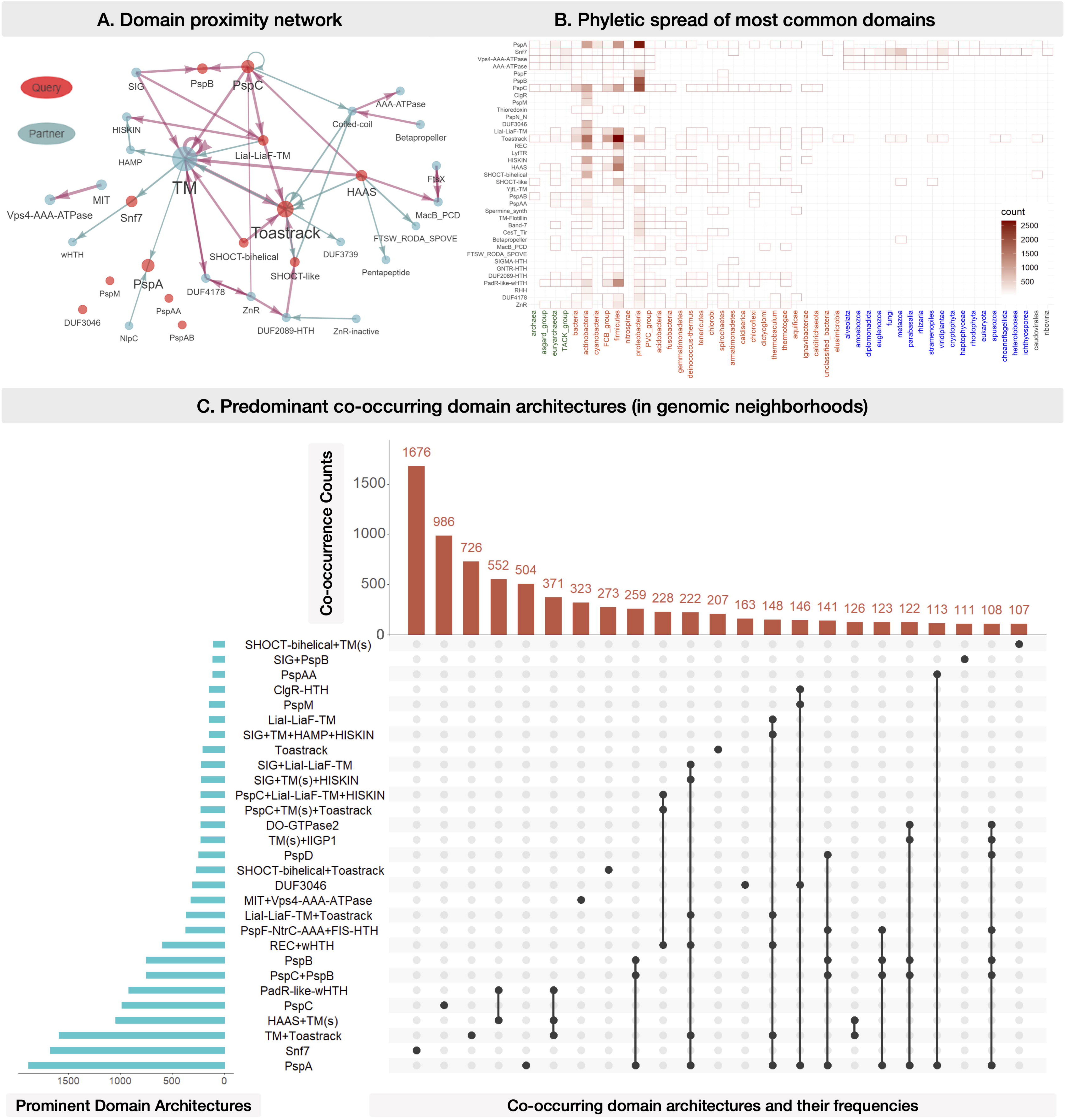

A
Unstressed
E. coli
Stressed
PspC
PspB
PspA
PspF
pspF
pspABCDE
Unstressed
B. subtilis
Stressed
LiaS
LiaF
LiaI
LiaH
LiaG
LiaR
liaIH
liaGFSR
Unstressed
M. tuberculosis
Stressed
MprB
MprA
SigE
ClgR
PspM
PspN
clgR
pspAMN
B
E. coli
pspF
FIS-HTH
PspF-NtrC-AAA
pspA
pspB
pspC
M. tuberculosis
ClgR-HTH
PspM
PspN_N
DUF3046
clgR
pspA
pspM
pspN
B. subtilis
LiaI-LiaF-TM
SIG
Toastrack
TM
HAMP
HISKIN
REC
wHTH
liaI
liaH
liaG
liaF
liaS
liaR
C
PspA
Snf7
PspB
PspC
PspM
PspN
LiaI-LiaF-TM
Toastrack
Kingdoms
bacteria
eukaryota
archaea
Phyla
asgard group
crenarchaeota
euryarchaeota
acidobacteria
actinobacteria
armatimonadetes
bacteroidetes
chloroflexi
cyanobacteria
deinococcus-thermus
firmicutes
gemmatimonadetes
planctomycetes
viridiplantae
spirochaetes
tenericutes
thermotogae
verrucomicrobia
alveolata
amoebozoa
choanoflagellida
diplomonadida
euglenozoa
fungi
ichthyosporea
metazoa
parabasalia
rhizaria
rhodophyta
stramenopiles
proteobacteria


Cyanobacteria
Firmicutes
Eukaryotes
Actinobacteria
Proteobacteria
Eukaryotes
Archaea
PspA
PDB: 4WHE
Snf7
PDB: 5FD7

PspA: Domain Architectures
TM PspA
bacteria, archaea, eukaryota
PspA 2-4
cyanobacteria
BAG06017.1
Microcystis aeruginosa NIES843
NlpC PspA
cyanobacteria
AFZ52345.1
Cyanobacterium aponinum PCC 10605
PspA PspAA
actinobacteria
ACU53894.1
Acidimicrobium ferrooxidans DSM 10331
PspA: Genomic Contexts
With PspF, PspB, PspC and related-contexts
ANW39986.1
Escherichia coli
FIS-HTH PspF-NtrC-AAA PspA PspB PspC PspB SIG PspD SIG RHOD-CDC25 DO-GTPase2 TMs IIGP1-DUF697 PspF-NtrC-AAA FIS-HTH
AEN65073.1
Enterobacter soli
PspA
AHA65439.1
Shigella dysenteriae 1617
PspB 2-6
AOL22920.1
Erythrobacter litoralis
proteobacteria, spirochaetes, nitrospirae
ANQ40502.1
Gluconobacter oxydans
FIS-HTH PspF-NtrC-AAA PspA PspB SIG Toastrack SIG Toastrack
α-proteobacteria
AMQ41621.1
Aeromonas veronii
PspA PspC Phageshock_PspG
γ-proteobacteria
AEC01378.1
Sphaerochaeta coccoides DSM 17374
PspA PspC FIS-HTH PspF-NtrC-AAA PspA PspB-like PspC PspB SIG Toastrack
spirochaetes
ANL45788.1
Rhizobium phaseoli
PspA YiaAB
α-proteobacteria
With PspAA
AAM04874.1
Methanosarcina acetivorans C2A
PspA PspAA SIG TM(s) HISKIN REC wHTH
CAB51252.1
Streptomyces coelicolor A32
actinobacteria, archaea
ABW11964.1
Frankia sp EAN1pec
PspA PspAA TM(s) Metallo peptidase TM(s) PspAB
actinobacteria
AKB54760.1
Methanosarcina barkeri MS
PspA PspAA PspAB TM(s) Metallo peptidase TM(s)
archaea, chloroflexi, firmicutes, actinobacteria, proteobacteria
SIG SHOCT-bihelical TM(s) Metallo peptidase TM(s) PspAB PspA PspAA
AKX93460.1
Moorella thermoacetica
firmicutes
With ClgR, PspMN or Thioredoxin
AKK09009.1
Corynebacterium testudinoris
clgR-HTH PspA PspM PspN_N DUF3046
CCP45543.1
Mycobacterium tuberculosis H37Rv
actinobacteria
AKK09942.1
Corynebacterium testudinoris
Thioredoxin PspA
actinobacteria
With chaperones
ANH61663.1
Dokdonia donghaensis DSW1
CesT_Tir-DUF2170 PspA SIG TM(s) SIG Band-7 Coiled-coil Flot Betapropeller AAA-ATPase PhospholipaseD
FCB, proteobacteria, firmicutes
AAN56746.1
Shewanella oneidensis MR1
RHH DUF4178 PspA CesT_Tir-DUF2170 Spermine_synth TM(s) SIG YjfL-TM(s) Low-comp Ctha_1186 LipoSIG
γ-proteobacteria, bacteroidetes
CCG46123.1
Halobacillus halophilus DSM 2266
TM DUF4178 PspA SIG DUF4247 YjfL-TM(s) SIG TM(s) Spermine_synth
firmicutes
AAS43862.1
Bacillus cereus ATCC 10987
PspA YjfL-TM(s) SIG DUF4178 SIG DUF4247
firmicutes
BAB38581.1
Escherichia coli O157:H7 str Sakai
CesT_Tir-DUF2170 PspA SIG Ion_trans_2 DUF2491 YjfL-TM(s) SIG DUF1190 SpermGS-ATPgrasp
γ-proteobacteria
AEF02306.1
Alteromonas naphthalenivorans
ANL88610.1
Rhizobium phaseoli
PspA CesT_Tir-DUF2170
α-proteobacteria
AAK24333.1
Caulobacter vibrioides CB15
CesT_Tir-DUF2170 PspA SpermGS-ATPgrasp
α-proteobacteria
ANX09535.1
Bacillus subtilis subsp subtilis
PspA ZnR(s) TM SIG TPM_phosphatase TM Band-7 ZnR
firmicutes
SIG TPM_phosphatase TM Coiled-coil PspA NYN REC wHTH SIG TM HAMP HISKIN
ALD73647.1
Trueperella pyogenes
AKL67689.1
Streptomyces sp. Mg1
actinobacteria, firmicutes, thermotogae
With two-component systems
ANX06812.1
Bacillus subtilis subsp subtilis
SIG LiaI-LiaF-TM PspA SIG Toastrack LiaI-LiaF-TM Toastrack SIG TM HAMP HISKIN REC wHTH
firmicutes
APB74393.1
Paenibacillus polymyxa
LiaI-LiaF-TM PspA PspC Coiled-coil PspA SIG TM Toastrack SIG TM(s) HISKIN REC wHTH
firmicutes
Cyanobacterial and related contexts
AFZ14666.1
Crinalium epipsammum PCC 9333
PspA PspA Thioredoxin
cyanobacteria
BAG04642.1
Microcystis aeruginosa NIES843
PspA SIG TPR PBPI
cyanobacteria
AIK12556.1
Pectobacterium atrosepticum
SIG TPR TM DUF5401
γ-proteobacteria

TM Toastrack
PspC
MIT Vps4-AAA-ATPase
HAAS TM(s)
Toastrack
SHOCT-bihelical Toastrack
PspB
PspC PspB
LiaI-LiaF-TM Toastrack
DUF3046
SIG PspB
PspC LiaI-LiaF-TM HISKIN
PspC TM(s)
PspC TM(s) Toastrack
ZnR DUF2089-HTH SHOCT-like
TM DUF4178
HAAS MACB_PCD FtsX MACB_PCD FtsX
Betapropeller Coiled-coil AAA-ATPase
SHOCT-like Toastrack
PspAA
PspAB
SIG TM Toastrack
PspM
SIG TM(s) Toastrack
SHOCT-bihelical TM(s)
LiaI-LiaF-TM
SIG LiaI-LiaF-TM Toastrack
SHOCT-like X
DUF4178
SHOCT-bihelical TM
DUF2089-HTH SHOCT-like
SIG LiaI-LiaF-TM
ZnR DUF4178 ZnR DUF4178 TM(s)
HAAS PspC LiaI-LiaF-TM Toastrack
HAAS FTSW_RODA_SPOVE
HAAS Toastrack
SHOCT-like
SHOCT-like Coiled-coil Toastrack
SIG TM HAMP HISKIN
Lineage
Viruses
Eukaryota
Bacteria
Bac>thermotogae
Bac>thermobaculum
Bac>tenericutes
Bac>spirochaetes
Bac>PVC_group
Bac>proteobacteria
Bac>nitrospirae
Bac>ignavibacteriae
Bac>gemmatimonadetes
Bac>fusobacteria
Bac>firmicutes
Bac>FCB_group
Bac>elusimicrobia
Bac>dictyoglomi
Bac>deinococcus-thermus
Bac>cyanobacteria
Bac>chloroflexi
Bac>chlorobi
Bac>calditrichaeota
Bac>caldiserica
Bac>armatimonadetes
Bac>aquificae
Bac>actinobacteria
Bac>acidobacteria
Archaea
Arc>TACK_group
Arc>euryarchaeota
Arc>asgard_group
0
500
1000
1500
2000
2500
Domain Architectures
Number of Occurrences per Lineage

With SHOCT-bihelical and SHOCT-like
SIG
TM(s)
SIG
TM(s)
HISKIN
REC
wHTH
SHOCT-bihelical
Toastrack
CCP43715.1
Mycobacterium tuberculosis H37Rv
actinobacteria, proteobacteria, spirochaetes, others
TM(s)
TM(s)
ABC-ATPase
SHOCT-bihelical
Toastrack
CAB88834.1
Streptomyces coelicolor A32
actinobacteria
CAB15517.1
Bacillus subtilis subsp subtilis str 168
yvlA-SIG
TM(s)
SHOCT-like
Toastrack
PspC
Toastrack
PspC
SIG
TM(s)
firmicutes
AQL53774.1
Enterococcus faecalis
SHOCT-like
Coiled-coil
Toastrack
PspC
(2)
SIG
TM(s)
BAI87136.2
Bacillus subtilis subsp natto BEST195
firmicutes
TM(s)
With PadR-like-wHTH (PadR-wHTH) and HAAS
AAO82145.1
Enterococcus faecalis V583
CAC98679.1
Listeria monocytogenes EGDe
AIR09900.1
Lactobacillus salivarius
PadR-wHTH
HAAS
TM(s)
(2)
SIG
Toastrack
PspC
firmicutes, proteobacteria, spirochaetes, others
CAC05954.1
Streptomyces coelicolor A32
Toastrack
PadR-wHTH
actinobacteria
AOW09537.1
Flavobacterium gilvum
PadR-wHTH
HAAS
PspC
LiaI-LiaF-TM
Toastrack
TM
Toastrack
SIG
Toastrack
bacteroidetes
PadR-wHTH
HAAS
HAAS
Toastrack
CAW95595.1
Streptococcus equi subsp equi 4047
firmicutes
AAO82145.1
Enterococcus faecalis V583
PadR-wHTH
HAAS
TM(s)
Toastrack
TM
AOJ43666.1
Burkholderia thailandensis
proteobacteria
PadR-wHTH
HAAS
PspC(s)
LiaI-LiaF-TM
Toastrack(s)
AKD55949.1
Spirosoma radiotolerans
bacteroidetes
PadR-wHTH
HAAS
MACB_PCD
FTSX
MACB_PCD
FTSX
ACO32024.1
Acidobacterium capsulatum ATCC 511
acidobacteria, gemmatimonadetes
AOH56696.1
Bacillus muralis
PadR-wHTH
HAAS
Pentapeptide
firmicutes
SIGMA-HTH
PadR-wHTH
HAAS
FTSW_RODA_SPOVE
CAC98500.1
Listeria monocytogenes EGDe
firmicutes
With Lia, two-component systems, other transcription factors
AJH41968.1
Bacillus anthracis
CAB15300.1
Bacillus subtilis subsp subtilis str 168
SIG
LiaI-LiaF-TM
PspA
SIG
Toastrack
LiaI-LiaF-TM
Toastrack
SIG
TM(s)
HAMP
HISKIN
REC
wHTH
SIG
TM(s)
firmicutes, bacteroidetes, archaea, others
CAL82154.1
Clostridium botulinum A str ATCC 3502
LytTR
LiaI-LiaF-TM
Toastrack
firmicutes
wHTH
REC
HISKIN
LiaI-LiaF-TM
PspC
PspC
TM(s)
Toastrack
AIJ13866.1
Streptomyces lividans TK24
actinobacteria
ACU05652.1
Pedobacter heparinus DSM 2366
PspC
TM(s)
LiaI-LiaF-TM
Toastrack
firmicutes, bacteroidetes
SIGMA-HTH
AS-ZF
TM
TM
LiaI-LiaF-TM
TM
SIG
LiaI-LiaF-TM
Toastrack
AEU34960.1
Granulicella mallensis MP5ACTX8
acidobacteria
ABM00816.1
Shewanella amazonensis SB2B
SIGMA-HTH
TM
TPR
SIG
Toastrack
TM
Toastrack
ANH61515.1
Dokdonia donghaensis DSW1
proteobacteria, bacteroidetes
SIG
SIG
TM(s)
SIG
TM(s)
HISKIN
REC
wHTH
SHOCT-bihelical
Toastrack
CRY81352.1
Nocardia farcinica
actinobacteria
ALR80200.1
Streptococcus salivarius
REC
wHTH
ABC-ATPase
SIG
TM(s)
Toastrack
SIG
TM(s)
Toastrack
firmicutes
APB74396.1
Paenibacillus polymyxa
ABD31150.1
Staphylococcus aureus subsp aureus
LiaI-LiaF-TM
PspA
PspC
Coiled-coil
PspA
SIG
TM(s)
Toastrack
SIG
TM
HISKIN
REC
wHTH
firmicutes
AAM36414.1
Xanthomonas citri pv citri str 306
GNTR-HTH
ABC-ATPase
TM(s)
SIG
Toastrack
SIG
DUF2884
AJI31452.1
Bacillus cereus E33L
proteobacteria, firmicutes
AGT82698.1
Amycolatopsis mediterranei RB
RHH
Toastrack
ABC-ATPase
TM(s)
actinobacteria
AFY83227.1
Oscillatoria acuminata PCC 6304
SIG
Toastrack
CASPASE
cyanobacteria



A. Domain proximity network

B. Phyletic spread of most common domains

C. Predominant co-occurring domain architectures (in genomic neighborhoods)