The Phage-shock-protein Envelope Stress Response: Discovery of Novel Partners & Evolutionary History

Janani Ravi[1,2]*, Vivek Anantharaman[3], Samuel Zorn Chen[1], Pratik Datta[2], L Aravind[3*], Maria Laura Gennaro[2]*.
[1]Pathobiology and Diagnostic Investigation, Michigan State University, East Lansing, MI, USA; [2]Public Health Research Institute, Rutgers University, Newark, NJ, USA; [3]National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD, USA. *Corresponding authors. janani@msu.edu; aravind@nih.gov; marila.gennaro@rutgers.edu

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

Abstract

The phage shock protein (PSP) systems orchestrate a conserved bacterial stress response function by stabilizing the cell membrane and protecting bacteria from envelope stress. However, the full repertoire of PSP components remains poorly characterized. We combined comparative genomics and protein sequence-structure-function analyses to systematically identify sequence homologs, phyletic patterns, domain architectures, and gene neighborhoods to trace the evolution of PSP components across the tree of life. We showed that different bacterial clades possess distinct PSP systems. This integrative approach traced the universal nature and origin of PspA/Snf7 (Psp/ESCRT systems) to the Last Universal Common Ancestor. We identified novel partners of the PSP system: one is the Toastrack domain, which likely facilitates assembling diverse sub-membrane stress-sensing and signaling complexes; another is the HAAS–PadR-like transcription factor system. We showed that PspA/Snf7 proteins have evolved associations with ATP-dependent, CesT/Tir-like chaperones, and Band-7 domain proteins to mediate sub-membrane dynamics. Our work also uncovered links between the PSP components and diverse SHOCT-like domains, suggesting a role in assembling membrane-associated complexes of proteins with disparate biochemical functions. Tracing the evolution of Psp cognate proteins provides new insights into the functions of the system and helps predict previously uncharacterized, often lineage-specific, membrane-dynamics and stress response systems. The conservation of PSP systems across bacterial phyla emphasizes the importance of this stress response system in prokaryotes, while its modularity in various lineages indicates adaptation to lineage-specific cell-envelope structures, lifestyles, and adaptation mechanisms. The results can be accessed at https://jravilab.shinyapps.io/psp-evolution.

Keywords

Phage shock protein response | Psp | Envelope stress response | Phylogeny and evolution | Domain architectures | Genomic contexts and neighborhoods

Introduction

Membranes are complex dynamical structures enclosing all living cells. They typically comprise lipid bilayers and other non-bilayer lipids as well as proteins [1]. The membrane and its associated elements partake in a variety of critical dynamic processes such as membrane biogenesis, cell division, maintenance of cell shape, transport of small molecules, cell-cell and cell-environment signaling, maintenance of proton motive force (PMF), and other structural processes involving cytoskeletal proteins [2]. Even though the biochemical composition and structure of the membranes and associated lipids and proteins can vary widely, most membrane functions are conserved across the tree of life [1]. Cellular membranes are continually changing to preserve structure and function in response to changing extracellular signals, particularly in unicellular organisms that are subjected to various types of stress originating from the external environment. Indeed, extra-cytoplasmic envelope stress responses are common virulence mechanisms employed by bacterial pathogens to survive to host insult [3–5]. Moreover, antibiotics and antimicrobial peptides can trigger the cell envelope stress responses [6,7]. The envelope integrity might also be compromised by lytic phage infection. Since membrane damage can ultimately result in cell death, a specialized suite of lipids and proteins participate in sensing and responding to this stress [8,9], initiating a chain of events to restore membrane function. Among the mechanisms involved in maintaining the biophysical properties of the cell membrane is the ESCRT system in eukaryotes and the Phage Shock Protein (PSP) system in bacteria [2,10–13]. The multi-protein PSP system constitutes a ubiquitous envelope stress response machinery found across bacterial phyla that comprises both conserved and lineage-specific components expressing conserved stress response functions [1,2,10,11,13–16,16–21]. Indeed, applying similar solutions to commonly encountered problems such as environmental stress via phenotypic convergence, even if not paired with genotypic convergence, is common [22]. To fully understand the molecular basis of membrane maintenance mechanisms, it is essential to map out how systems such as PSP have evolved in the course of the history of life.

The literature reports that the PSP system is centered on the protein PspA, a peripheral membrane associated with the intracellular face of the cell membrane. Phylogenetic analysis has revealed that PspA is homologous to Vipp1, a plant protein of cyanobacterial origin involved in the vesicular transport and storage of lipids, thylakoid biogenesis, and membrane-shaping [16,20,23–28]. Vipp1 bears similarity to other bacterial and eukaryotic protein families involved in secretory pathways, protein sorting and transport, and stabilization of bent membranes. Structurally, both PspA and Vipp1 contain the *PspA_IM30* domain with a predominantly coiled-coil structure [14,16,19,24,25,29–32] [**Table S1**]. The N-terminal amphipathic helices associate with membranes to sense stored curvature elastic stress [13,31]. The *PspA_IM30* domain is found in bacteria, photosynthetic eukaryotes, and archaea [17,18,20,21,24,33–36] [**Fig. 1**]. Both Vipp1 and PspA are present in the photosynthetic cyanobacteria [16,24,37], hinting at the domain's origins and phylogenetic migration over to eukaryotes. Of the two, only PspA retains the stress sensing role [7,15,18,38–40], while Vipp1 is involved in thylakoid biogenesis [16,25,26]. For both proteins, the primary role is to preserve envelope function in response to a wide range of surface stresses via membrane fusion [16] and prevent leakage of protons through the membrane [2,15]. These perturbations include changes in

membrane potential, Δ*pH* [2,15] and PMF [41–44], (l)antibiotic stress [7], heat/osmotic shock [45] and mechanical stress via stored curvature elastic stress [14,46–48], among others [45].

The bacterial PSP system includes several other proteins, typically transcriptional regulators and other integral membrane proteins that function as partners of PspA. The three best described PSP systems span the dominant bacterial lineages, namely, proteobacteria (*pspF||ABCDE-G* [19,34,44,49–54]), firmicutes (*liaIHGFSR* [21,35,40,55–58]), and actinobacteria (*clgRpspAMN* [17,18,36] and Datta *et al.*, unpublished) [**Fig. 1**]. These cognate members serve distinct roles such as transcriptional regulation (such as PspF, ClgR, the three-component LiaFSR system) or membrane tethering and relocation (PspB, PspC, PspM, LiaI, LiaF). Some of the partner proteins such as PspC/PspB [42,52,59–64] or the LiaRS two-component system [21,40,55,56,65] have been implicated in direct stress-sensing functions independent of PspA. However, the classical PSP system initially discovered in Proteobacteria [19,29] is not universal, *i.e.*, the system is not conserved across all bacterial/archaeal clades. Despite the variable genomic contexts, PspA seems to have evolved always to preserve the cell membrane. Current studies have mostly focused on lineage-specific PSP systems (*e.g.,* Proteobacteria, Firmicutes, Actinobacteria, Plants [3,17,18,20,24,25,33,36,44,66,67]). It remains unclear how PspA has evolved to function in membrane homeostasis across species and the nature of the most likely cognate partners that help PspA achieve this pivotal role.

In the present work, we conducted an evolutionary analysis across all the major kingdoms of life [68,69] to delineate all possible occurrences of PspA and its cognate partners, their associated genomic contexts, and phyletic spreads across the bacterial and archaeal kingdoms. To do so, we utilized a generalized and global approach in which we resolved Psp proteins into their constituent domains and used each of the individual domains as queries for all our phylogenetic analyses. This domain-centric characterization approach is best suited to help address the following questions regarding the evolutionary and functional significance of the PSP system: 1) Which are the most distinct and commonly occurring themes in terms of a) domain architectures and b) conserved gene neighborhoods that house the Psp proteins? 2) Have the Psp proteins evolved functions beyond those discussed above? 3) Are there any other common themes that emerge from the characterization of the phyletic patterns of PSP systems? 4) Finally, do the gene neighborhoods and phyletic patterns of the Psp components imply greater independence of the individual subsystems than initially anticipated? The resulting findings presented here help identify novel partners of the PSP envelope stress response system and throw light on both its evolution and function.

Results and Discussion

To understand the evolution of the PSP system and its ability to function across phyla, we needed to reconcile the universality of PspA with the distinct genomic neighborhoods containing PspA [17,33,66] in each of the distinct bacterial, archaeal, and eukaryotic lineages. The first step towards this goal was to identify and characterize all close and remote homologs of PspA and cognate partner domains across the tree of life.

To locate and stratify all Psp homologs, we started with the previously characterized Psp proteins from the best-studied operons *pspF||ABCDE-G* (from *E. coli*), *liaIHGFSR* (from *B. subtilis*), and *clgRpspAMN* (from *M. tuberculosis*), and analyzed their phyletic spread [**Fig. 1**]. To ensure an exhaustive search and identification of all related Psp proteins, including remote homologs, we first resolved these Psp proteins into their constituent domains and used each of the individual domains as queries for all our phylogenetic analyses. This approach allows us to find homologies beyond those pertaining to full-length proteins only. Homology searches for each constituent domain were performed across the tree of life (all 6500+ completed genomes; see *Methods*). We used a combination of protein domain and orthology databases, iterative homology searches, and multiple sequence alignment to detect domains, signal peptides, transmembrane (TM) regions to help construct the domain architecture of each of the query Psp proteins [see *Methods*]. Due to the ubiquitous presence of transcription factor Helix-turn-helix (HTH)-type domains (*e.g.,* PspF, ClgR) and Histidine kinase-receiver domain two-component systems in bacterial phyla (*e.g.,* LiaRS; **Fig. 1A**), we did not perform dedicated searches with these domains; instead, we noted their occurrence in the predicted operons alongside the core Psp genes in order to identify transcriptional regulation partners. Where required, as in the case of the Histidine Kinase, we identified the closest orthologs to study the full extent of the diversity.

Our results are organized as follows. First, we present the delineation of the core Psp proteins in proteobacteria, actinobacteria, and firmicutes, and describe their underlying domain architectures and their phyletic spreads [**Fig. 1**]. Then, we outline the conservation and evolution of the central protein, PspA, the systematic identification of PspA-containing species (spanning all the completed genomes across the bacterial, archaeal, and eukaryotic kingdoms) [**Fig. 2**], followed by the molecular characterization of the homologs in terms of domain architecture, genomic context, and phyletic spread [**Fig. 3**]. Finally, we report the findings from our in-depth analysis of domain architecture [**Fig. 4**], genomic context [**Fig. 5**], and phyletic spread of each of the Psp cognate partners to identify novel components (domains and genomic neighborhoods) of the PSP system across the tree of life [**Fig. 4, 5**]. We then reconciled all our findings in the form of a proximity network of domains with respect to PspA along with their phyletic spread and frequencies of occurrence across the tree of life [**Fig. 6**]. All our findings (data, results, and visual summaries) are available in an interactive and queryable web application https://jravilab.shinyapps.io/psp-evolution.

The known components of Psp and their domain architectures

The PSP system in proteobacteria

We initiated our analysis with the Psp operon first identified in *E. coli* and then in other proteobacteria [19,20,29,70–72]. We resolved the core members of the operon, PspF||ABC, into their domain architectures. PspA, PspB, and PspC each comprise their namesake domains, *PspA_IM30* (*IM30*, standing for inner membrane 30KDa domain, henceforth called PspA), PspB, and PspC, respectively, and each domain spans almost the entire protein length [**Fig. 1A**; **Table S1**]. In line with previous studies [51], we note that PspF, the transcriptional regulator of the Psp operon in *E. coli*, contains an enhancer-binding NtrC-AAA (a specialized AAA-ATPase) domain in addition to the Fis-type HTH (helix-turn-helix) DNA-binding domain [**Fig. 1B**].

The PSP system in actinobacteria

The protein central to the discovery of the Psp operon in *Mycobacterium tuberculosis* (actinobacteria) is PspA (encoded by *rv2744c*) [18]. We confirmed that PspA contains the PspA domain, albeit with <30% similarity to the *E. coli* homolog. In contrast to the PspF transcriptional regulator in *E. coli*, we observed that the *M. tuberculosis* operon includes a transcription factor, ClgR, with a cro-like HTH DNA-binding domain, which is unique to actinobacteria [**Fig. 1B**]. Moreover, the *M. tuberculosis* integral membrane protein, Rv2743c, is distinct from its *E. coli* counterparts, PspC and PspB [**Fig. 1B**]. The last member of the four-protein actinobacterial operon, Rv2742c, contains a domain of unknown function, DUF3046 [**Fig. 1B**; **Table S1**]. Despite its requirement for the full activity of the PSP system [18], the molecular role of the Rv2742c protein remains uncharacterized. These findings are in line with our previous studies [17,18,36]. To remain consistent with the nomenclature of Psp cognate proteins, Rv2743c and Rv2742c were renamed as PspM and PspN, respectively [Datta *et al.,* unpublished].

The PSP system in firmicutes

When we examined the domain architectures of the well-studied Lia operon in *B. subtilis*, we found that, in line with recent studies [21,56,65], LiaI is a small TM protein; LiaF is a protein with N-terminal TM helices (DUF2157) and a C-terminal globular domain (DUF2154), and LiaG contains another domain of unknown function (DUF4097 that is similar to DUF2154) [**Fig. 1B**]. Our analysis helped clarify relationships of these aforesaid domains and correctly define two new domains of the PSP system: i) the **LiaI-LiaF-TM** sensory domain encompassing 4 TM domains (4TM) and DUF2157, and ii) **Toastrack** encompassing the DUF4097/DUF2154/DUF2807 domains (see the section below for details on our newly defined domain nomenclatures, LiaI-LiaF-TM and Toastrack) [**Fig. 1B**; **Table S1**]. As documented previously [21,40], we found that, in this firmicute, the transcriptional response is expressed by the two-component system, LiaRS, which comprises of the LiaR protein with a receiver domain and a winged HTH (wHTH) DNA-binding domain (REC–wHTH) and a sensor kinase LiaS with a TM region, intracellular HAMP and a Histidine kinase signaling domain (HAMP–HISKIN) [73] [**Fig. 1B**; **Tables S1, S2**].

In summary, the above data show that, in addition to the transcriptional regulators and previously described domains, such as PspA, PspB, PspC in proteobacteria, we defined novel domains, such

as LiaI-LiaF-TM and Toastrack, in firmicutes [**Fig. 1B**; **Table S1**]. We next focused on the presence of these proteins across the tree of life.

Psp components across the tree of life

PspA

We started with a detailed analysis of the phyletic spread of PspA. Consistent with previous PspA evolutionary studies, we found that PspA, containing the PspA domain (previously PspA_IM30; spanning almost the entire length of the *E. coli*, *B. subtilis*, and *M. tuberculosis* PspA proteins; **Table S1**), is present in most major bacterial and archaeal lineages in one or more copies [**Fig. 1C**]. We observed that, within eukaryota, only the SAR group (the clade that includes stramenopiles, alveolates, and rhizaria) and archaeplastids (including plants and algae) carry PspA homologs [**Fig. 1C**]. We then performed a more detailed analysis of the PspA homologs in plants (Vipp1) using homology searches with multiple distinct and distant PspA proteins. This analysis revealed that the PspA_IM30 (henceforth referred to as PspA) domain is homologous to the eukaryotic Snf7 superfamily [**Fig. 1B**], which is part of the membrane-remodeling ESCRT system in eukaryotes [11,74–78]. Further, iterative searches with PspA/Vipp1 and Snf7 recovered each other, implying that these proteins are part of the same superfamily. In addition, iterative searches with the Snf7 domain recovered several non-plant eukaryotes and archaeal homologs [**Fig. 1B**]. Together, the PspA and Snf7 superfamilies are widely prevalent, as they extend to almost every phyletic lineage spanning all three kingdoms [**Fig. 1C**]. The PspA branch is conserved in bacteria with transfers to archaea and eukaryotes, while the Snf7 branch is conserved in archaea and eukaryotes with transfers to bacteria. This part of the analysis shows that PspA/Snf7 is the only 'universal protein' in the PSP system present in all three kingdoms: bacteria, archaea, and eukaryota [**Fig. 1B**; **Table S3**].

LiaI-LiaF-TM and Toastrack

Next, we traced the occurrence of LiaF, LiaG, and LiaI across the tree of life using their full-length protein sequences and their constituent domains, including our new domain definitions, LiaI-LiaF-TM and Toastrack. [**Fig. 1B**]. We used PSI-BLAST searches from the three sub-sequences of the full-length proteins, N-terminal TM region of LiaF, the C-terminus globular domain (DUF2154) of LiaF, and the globular domain in LiaG (DUF4097), followed by structure-informed sequence alignment ['MSA' tab in the webapp]. These analyses revealed that LiaI and LiaG bear remarkable similarities to the N-terminal TM and C-terminal globular regions of the LiaF protein, respectively. Indeed, we discovered that the globular domains of these LiaG–LiaF proteins are homologs of each other and that the profiles detected by Pfam in this region, DUF2154, DUF4097, and DUF2807, are unified into a single domain that has a single-stranded right-handed beta-helix like structure called Toastrack (PDB: 4QRK; **Table S1**; 'MSA' tab in the webapp). Therefore, we collectively refer to these three "domains of unknown function" as the '**Toastrack**' domain [**Fig. 1A**; **Table S1**]. Likewise, the homology between the 4TM (four TM) regions of LiaI and the N-terminal domain DUF2157 of LiaF led us to rename the 4TM region (previously referred to as Toastrack_N, Pfam: PF17115) to be '**LiaI-LiaF-TM**' [**Fig. 1B**; **Table S1**]. Thus, the results of our analyses define two new domains: LiaI-LiaF-TM and Toastrack. Toastrack-containing proteins are pan-bacterial with transfers to archaea and eukaryotes [**Fig. 1B**]. We also found a few metazoan, fungal, and SAR-group eukaryotic

genomes that carry this globular *Toastrack* domain [**Fig. 1B**]. The Toastrack domain is frequently coupled to the LiaI-LiaF-TM (*e.g.,* CAB15300.1, Bacillus), PspC (ABY34522.1, Chloroflexus), or a combination of both [**Fig. 4**; **Table S4**; ACV77657.1, *Nakamurella*]. LiaI-LiaF-TM is found mainly in firmicutes, actinobacteria, bacteroidetes, unclassified bacteria, thermotogae, acidobacteria, chloroflexi, with sporadic occurrences in other clades and transfers to euryarchaeota [**Fig. 1B**]. The LiaI-LiaF-TM shows further fusions to terminal B-box or the SHOCT-like domains (see below). Given that the LiaI-LiaF-TM domain associates in conserved gene-neighborhoods or domain architectures with different intracellular signaling modules, we propose that it serves as an integral-membrane sensor that transmits signals via these associated modules. We describe in detail the novel domain architectures and genomic contexts featuring these domains in the sections below.

PspC

PspC, an integral membrane protein first identified in the proteobacterial PSP system, is critical in sensing the membrane stress response and restoring envelope integrity [59,71,79]. Recent studies have shown that PspC may function independently of PspA in some bacterial species in response to membrane stressors such as secretin toxicity [59,61,71,79,80]. Our analysis showed that PspC has two TM helices, the first being a cryptic TM with an unusual residue composition. A conserved amino acid, R, is found between the two TM regions, suggesting a role in membrane-associated sensing of a stimulus ['MSA' tab in the webapp]. We found that the PspC domain is pan-bacterial and also present in a few archaeal clades [**Fig. 1C**]. We also identified several novel domain architectures and genomic contexts of PspC (described below). For example, in some firmicutes, PspC is fused to an additional C-terminal coiled-coil domain, and these variants always occur in operons coding for two flanking PspA-genes [see **Fig. 4, 5** in later sections on *PspC domain architectures and genomic neighborhoods*].

PspB

PspB is another integral membrane protein that is often part of the proteobacterial Psp operon and is implicated in stress sensing and virulence [71,80,81]. PspB has an N-terminal anchoring TM helix followed by an α-helical globular domain. We rarely found PspB in other bacterial lineages outside proteobacteria, but we discovered previously unrecognized divergent PspB fused to PspC [**Fig. 1B**; see the section below on *PspC domain architectures*].

PspM/PspN

We found no discernible homologs for these proteins (and constituent domains) outside the phylum actinobacteria [**Fig. 1C**], in accordance with our previous work [17,18,36].

PspM (encoded by *rv2743c*), the corynebacterial integral membrane partner, comprises two *TM* regions and no other distinct domain. Our phylogenetic analyses of the homologs revealed that PspM has minimum variation as shown by the multiple sequence alignment ['MSA' tab in the webapp], and a relatively narrow phyletic spread restricted to the genus Mycobacterium and its Corynebacterial neighbors (Rhodococcus, Nocardia, Actinopolysporales, Frankiales,

Micromonosporales, Nakamurellales, and Pseudonocardiales) [**Fig. 1B**], as per our previous findings [17,18,36].

The fourth member in the Mycobacterial operon, PspN (Rv2742c), contains a short domain at the C-terminus, DUF3046, and a yet uncharacterized N-terminal domain, which we now call PspN_N [**Table S1**]. We found that the DUF3046 domain is widely prevalent across actinobacteria but not in other phyla [**Fig. 1C**], as observed in our previous studies [17,18]. Moreover, the *M. tuberculosis* genome carries a second copy of DUF3046 (Rv2738c), located only four genes downstream of PspN (encoded by *rv2742c*). We found that DUF3046 is α-helical with highly conserved threonine and cysteine residues that might be required for its function (multiple sequence alignment in 'MSA' tab in the webapp).

To further characterize the DUF3046 homologs, we used nucleotide sequences rather than translated open reading frames (ORFs, followed by sequence alignment analysis [**Fig. 1B**]. We found that the DUF3046 domain, which is widespread across actinobacteria, is more similar to the short downstream protein, Rv2738c, than to the C-terminus of the fourth member of the *M. tuberculosis* operon, PspN (encoded by *rv2742c*). We thus infer that Rv2738c, rather than PspN, contains the ancestral copy of the DUF3046. The DUF3046 domain found as part of PspN is likely a duplicated copy of the domain that is translocated into the PspN ORF of mycobacteria, especially the *M. tuberculosis* complex. Moreover, unlike the pan-actinobacterial DUF3046, the N-terminal portion of PspN, 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* complex).

In summary, we have defined novel domains with multiple sequence alignments [**Fig. 1B**; webapp; **Table S1**] and delineated the phyletic spread of PspA and its cognate partner proteins/domains, PspBC, PspMN (DUF3046), LiaIFG (LiaI-LiaF-TM, Toastrack) [**Fig. 1C**]. While the membrane domains, PspB, PspM, LiaI-LiaF-TM, are highly lineage-specific (they are restricted to proteobacteria, actinobacteria, and firmicutes, respectively), the other two domains, PspC and Toastrack, are much more widely spread [**Fig. 1C**]. In addition to being present in a wide range of bacterial phyla, PspC is found in archaeal clades, and Toastrack is present in archaea and a few eukaryotic lineages [**Fig. 1C**]. The complete list of homologs and phyletic spreads are available under the 'Data' and 'Phylogeny' tabs in our interactive webapp. To further characterize the homologs of PspA, its cognate partner domains, and other novel domains found in the proximity of these PSP systems, we next carried out a detailed analysis of domain architecture, genomic context, and phyletic spread of each of these domains.

Evolution of PspA

We first delved deeper into our finding of the ancestral PspA/Snf7 superfamily with the help of structure-informed multiple sequence alignment and phylogenetic tree of homologs of the PspA/Snf7 superfamily ['MSA' tab in the webapp].

Identifying PspA+ and Snf7+ clades

PspA+: As noted above, to perform an extensive and inclusive search to identify all the close and remote homologs of PspA, we started with six distinct starting points as queries (PspA from *E. coli*, *M. tuberculosis*, two copies of *B. subtilis*, and representatives from cyanobacteria, viridiplantae (Vipp1)). We found that most bacterial clades contain PspA homologs [**Fig. 1B**] with a few instances of transfers to archaea. Among eukaryotes, only those containing plastids and two flagella (comprising the SAR/HA supergroup and excavata) have homologs with the PspA domain [**Fig. 2; 1B**; [82,83]].

Snf7+: Snf7, a predominantly archaeal and eukaryotic protein [**Fig. 1B**], is part of the ESCRT-III complex required for endosome-mediated trafficking via multivesicular body formation and sorting [11,74–78]. Examining the curated set of PspA-like proteins revealed a distinct cluster of proteins belonging to the eukaryotic Snf7 family among the remote PspA homologs [**Fig. 2**; see *Methods*; 'Data' tab of webapp]. We next explored the shared evolution of PspA and Snf7 homologs across the tree of life.

Tracing the phylogeny of PspA/Snf7 to the last common ancestor

To resolve the history and evolution of the PspA-Snf7 superfamily, we started with a comprehensive selection of PspA/Snf7 homologs with archetypical representatives from the distinct clades across the tree of life ['Data' tab of webapp; [68,69]] as well as different domain architectures [**Fig. 2**; detailed in the next section]. We also included PspA/Snf7 paralogs and incorporated any available structural information for PspA and Snf7 to better inform the sequence alignment, such as low complexity coiled-coils in both proteins (*E. coli* PspA (4WHE) and *S. cerevisiae* Snf7 (5FD7) [**Fig. 2A**; **Table S1** [84,85]]. With the representative PspA/Snf7 proteins spanning the three kingdoms, we performed a multiple sequence alignment with the selected ~450 PspA/Snf7 homologs ['MSA' tab in the webapp; see *Methods*]. Finally, we used this multiple sequence alignment to generate a PspA/Snf7 phylogenetic tree [**Fig. 2A**] to trace the evolutionary history of the superfamily. The multiple sequence alignment of PspA/Snf7 also highlights variation in position, length, and nature of the α-helices (*e.g.,* four helices are found in most bacterial and archaeal lineages [webapp]). Of note is a unique insertion of heptad repeats in actinobacteria, likely conferring a membrane- and partner-specific adaptation. We also found a C-terminal extension in a few cyanobacterial PspA homologs that we find to be more similar to the ancestral plant variant, Vipp1 [14,86].

To further investigate the phylogeny of PspA/Snf7, we used sequence alignment of a representative subset of species to construct a phylogenetic tree [**Fig. 2A**; see *Methods*]. The Snf7 homologs serve as the out-group in our PspA tree. The first striking finding is that several bacterial (including proteobacteria, actinobacteria, firmicutes) and archaeal clades, and their presumably ancestral cyanobacterial relatives [**Fig. 2A**], self-organize into distinct clusters due to their sequence-structural similarity and, presumably, the natural course of evolution. In addition to the clade-specific segregation of the PspA/Snf7 homologs, the principal observations from the tree (and the underlying sequence alignment) [**Fig. 2A**] are: i) the homologs from Actinobacteria, Firmicutes, and Proteobacteria form easily distinguishable clusters; ii) the cyanobacterial and eukaryotic homologs branch out (top left), in agreement with our earlier observation point to a specific-sequence

relationship between Vipp1 and the cyanobacterial PspA; and iii) the Snf7 domain-containing homologs from archaea and eukaryota form a well-defined branch separated by a long internal branch from the PspA containing branches.

In summary, homologs from the PspA and Snf7 superfamilies reveal that their shared helical membrane-stress response domain was inherited from the last universal common ancestor (LUCA). We have constructed a representative phylogenetic tree of PspA/Snf7 [**Fig. 2A**] based on hand-curated multiple sequence alignment (webapp) and characterized the PspA/Snf7 homologs in greater detail.

PspA: Novel architectures and neighborhoods

We leveraged our domain-level search to delineate the various domain architectures and genomic contexts of the PspA homologs in genomes from diverse clades.

Domain Architectures

Our comprehensive searches starting with eight representative PspA proteins revealed that most PspA homologs (>98%; ~2500 homologs; **Fig. 2B**) do not show much variation in their underlying domain architecture. In most lineages, PspA homologs only contain the characteristic PspA (PspA_IM30) domain without additional fusions [**Fig. 2B, 3**; **Table S3**]. The remaining small fraction of homologs (<2%) contain either repeated PspA domains or fusions with domains other than PspA [**Fig. 2B, 3**]. For example, cyanobacterial PspA homologs show some interesting variations: a few have dyads or triads of PspA, either as repeated domains within a polypeptide or a predicted operon with multiple copies of PspA-containing genes [**Fig. 3**; *e.g.,* BAG06017.1], while others carry an additional hydrolase domain of NlpC/P60 superfamily at the N-terminus that is predicted to catalyze the modification of phosphatidylcholine, thus altering membrane composition [**Fig. 3**; **Table S3**; AFZ52345.1; [87]]. We also find a novel fusion of PspA with **PspAA** in actinobacteria (ACU53894.1, Acidimicrobium; defined in the section on PspAA below). Similar to the PspA homologs, a search for the related superfamily, Snf7, revealed minimal variation in domain architecture, with occasional fusions (<5%) found only in eukaryotes [**Fig. 2B**]. Some actinobacteria, such as *Mycobacteroides abscessus*, have an Snf7 homolog (CAM62382.1, **Fig. 2B**) fused to an RND-family transporter member. The latter transports lipids and fatty acid and is flanked by two genes encoding the Mycobacterium-specific TM protein with a C-terminal Cysteine-rich domain [88].

Paralogs

We found that, in genomes that contain multiple copies of PspA, the paralogs do not maintain the same domain architecture and genomic context (discussed further below) [**Fig. 3**; **Table S3**; 'Phylogeny –> Paralog' tab of the webapp]. In cyanobacteria that carry multiple PspA paralogs, we observed that at least two of them occur as adjacent proteins in the genome [**Fig. 3**; **Table S3**; *e.g.,* BAG06015.1, BAG06016.1, BAG06017.1 in Microcystis]. This configuration is likely related to the scenario described above involving cyanobacterial dyads/triads with the variation that an insertion (or deletion) of intergenic sequence could have resulted in neighboring proteins rather than multi-PspA domain fusions (or vice-versa). Similar to previous studies [24,26], we observe that these neighboring

PspAs are part of two distinct clusters of homologs, one resembling the bacterial PspA and the other the eukaryotic (plant) Vipp1, suggesting a likely origin for the divergence in the PspA tree [**Fig. 3**]. In parallel to cyanobacterial operons containing adjoining PspA genes, a few archaeal and eukaryotic species also carry Snf7 gene clusters (CBY21170.1*; Oikopleura dioica*; **Table S3**), suggestive of tandem gene duplication and a conventional role in oligomerization and membrane stabilization as in the eukaryotic ESCRT systems [85]. Further analysis of PspA/Snf7 paralogs, including their likely evolution (gene duplication *vs.* horizontal gene transfer inferred from domain architectures and genomic contexts), can be found in the webapp ('Phylogeny' –> 'Paralog').

The occurrences of PspA and Snf7 in conserved gene-neighborhoods, albeit limited in their repertoire, are prevalent in all three kingdoms of life, sometimes containing multiple copies within a genome [**Fig. 2B, 3**; **Table S3**]. Interestingly, we observed a few co-occurrences of PspA and Snf7 within the same genomic neighborhood (AKJ06548.1 in deltaproteobacteria, CBH24266.1 in bacteroidetes with a third gene also containing a coiled-coil structure reminiscent of the PspA coiled-coils; **Table S3**). The ubiquity of the PspA/Snf7 superfamily indicates its inherent utility as the maintainer of envelope integrity.

Novel variations of known genomic contexts

We next explored the gene neighborhood for each of the Psp members [**Fig. 3**; **Table S3**]. The major contextual themes are summarized below. The full list of genomic contexts and their phyletic spread are available in the webapp ('Data' and 'Genomic Contexts' tabs, PspA/Snf7 selection).

PspFABC operon

We found that PspA, PspB, and PspC are found in an operon structure in alpha-, delta-, and gammaproteobacteria, and a few spirochaetes and nitrospirae [**Fig. 3**; **Table S3**]. We also found a few variations to this theme. In a few species, PspC is fused to a divergent C-terminal PspB in addition to a solo PspB in the operon [**Fig. 3**; **Table S3**; *e.g.,* ANW39986.1, *E. coli*], while others carry multiple PspB copies in the operon [**Fig. 3**; **Table S3**; *e.g.,* AOL22920.1, *Erythrobacter litoralis*]. PspD occurs along with this operon only in gammaproteobacteria [**Fig. 3**; **Table S3**; *e.g.,* ANW39986.1, *E. coli*]. The transcription regulator PspF (NtrC-AAA and HTH) is encoded divergently on the opposite strand but shares a common promoter with the PspABC operon in most gammaproteobacteria or with just the PspA gene in the remaining organisms that contain it. The entire operon also exists in a few spirochaetes (*e.g.,* Spico_0138), with additional variations involving PspB and PspC fusions and Toastrack containing proteins.

We also found that few operons in gammaproteobacteria (*e.g.,* Entas_2342 AEN65073.1) encode ligand-binding ACT and PAS domains [89], followed by the NtrC-AAA^{+}-domain [**Fig. 3**; **Table S3**]. In these occurrences, genes for proteins of the DO-GTPase family, predicted to play a role in membrane-related stresses [90], have been inserted between the NtrC-AAA+ protein and PspABCD. A few of these contain an additional canonical PspF. In some alphaproteobacteria, the PspC in PspFABC has been replaced by multiple Toastrack-containing proteins [**Fig. 3**; **Table S3**; *e.g.,* ANQ40502.1, *Gluconobacter oxydans*].

Associations with Vps4 and other classical AAA⁺-ATPases

One or more copies of an Snf7 gene (*e.g.,* OLS27540.1; **Table S3**) and a gene for the **VPS4-like AAA⁺-ATPase** (with an N-terminal **MIT** domain and C-terminal oligomerization domain; **Table S2**) are known to occur together in archaea; they define the core of an ESCRT complex [91]. However, we observed some diversity between different archaeal lineages. For example, the Asgardarchaeota contain a genomic context that is most similar to eukaryotes. This archaeal context is composed of the Vps4 AAA⁺-ATPase and the Snf7 genes along with an ESCRT-II gene that codes for a protein with multiple winged helix-turn-helix (wHTH) domains [92]. In crenarchaeota, Snf7 and the Vps4 AAA+-ATPase are encoded in a distinct three-gene operon, which contains a gene coding for a CdvA-like coiled-coil protein with an N-terminal PRC-barrel domain implicated in archaeal cell-division [93]. In this case, the Snf7 domain is fused to a C-terminal wHTH domain, which might play a role equivalent to the ESCRT-II wHTH domain. These operons may be further extended with additional copies of Snf7 genes and other genes coding for a TM protein and an ABC ATPase. We also observed that a related VPS4-like AAA⁺-ATPase was transferred from archaea to bacteria and is found in cyanobacteria, bacteroidetes, verrucomicrobia, nitrospirae, and planctomycetes (*e.g.,* ACB74714.1 *Opitutus terrae*; **Table S3**). In these operons, the Snf7 gene is displaced by an unrelated gene coding for a larger protein with TPR repeats followed by a 6TM domain, again suggesting a membrane-proximal complex.

Our analysis also showed that the bacterial PspA (*e.g.,* AEY64321.1, Clostridium; **Table S3**) might occur with a distinct AAA⁺-ATPase in various bacterial clades. The resulting protein (*e.g.,* AEY64320.1, Clostridium) has two AAA⁺-ATPase domains (*e.g.,* CKH37208.1, Mycolicibacterium) in the same polypeptide, with the N-terminal version being inactive. This gene dyad also occurs with either a previously unidentified membrane-anchored protein with a divergent Snf7 domain (OGG56892.1; **Table S3**) and other coiled-coil or α-helical domain-containing proteins. Both PspA and the membrane-associated Snf7, along with the AAA⁺-ATPase, may occur in longer operons with other genes coding for an ABC-ATPase, an ABC TM permease, and a solute-binding protein with PBPB and OmpA domains (*e.g.,* OGG56892.1; **Table S3**).

Membrane dynamics with PspA and AAA⁺-ATPase

Snf7 and VPS4 have previously been shown to be involved in the ESCRT-III-mediated membrane remodeling in archaea [74–76,78,85]. Similar to the Snf7 association with AAA⁺-ATPase, we found an operon containing PspA and AAA⁺-ATPase in bacteria, suggesting an involvement of the latter two proteins in bacterial membrane dynamics [**Fig. 3**; **Table S3**]. In bacteria, the PspA–AAA⁺-ATPase module is also sporadically part of a novel complex that contains both membrane-linked and cytoplasmic PspA/Snf7 homologs and additional transporter components. This novel complex suggests that, in response to stress: i) the PBPB-OmpA domain proteins interact with extracellular peptidoglycan, and ii) a ligand bound by the PbpB domain is transported by the ABC transporter. Further, we predict that this transporter complex, in turn, interacts with the PspA–AAA⁺-ATPase-dependent membrane remodeling proteins.

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

We found two novel overlapping genomic associations across various bacteria linking PspA with the Band-7 domain and CesT/Tir-like chaperone domain proteins [**Fig. 3**; **Table S2**; 'MSA' tab in the webapp]. Band-7 has previously been implicated in a chaperone-like role in binding peptides as part of the assembly of macromolecular complexes in different subcellular systems such as RNA repair [94]. Similarly, 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 [95]. We established by profile-profile searches that a previously uncharacterized protein encoded by genes linked to the *yjfJ* family PspA genes (*e.g.,* ANH61663.1, Dokdonia) is a novel member of the **CesT/Tir superfamily** [**Fig. 3**; **Table S3**; ANH61662.1, Dokdonia]. Similarly, we also observed that the proteobacterial proteins in the neighborhood of PspA [BAB38581.1, *E. coli*, **Fig. 3**; **Table S3**], typified by YjfI from *E. coli* (*e.g.,* BAB38580.1, DUF2170 in Pfam; **Table S2**), also contained a CesT/Tir superfamily domain [31]. We briefly describe below the gene neighborhoods linking proteins with CesT/Tir and Band-7 domains to PspA.

One class of conserved gene-neighborhoods is centered on a core two-gene dyad, comprising a CesT/Tir gene followed by a PspA gene [**Table S2**]. This two-gene core occurs either as a standalone unit or is further elaborated to give rise to two distinct types of larger operons. The first prominent elaboration combines the CesT/Tir–PspA dyad (ANH61663.1, Dokdonia) with i) a gene coding for a membrane-associated protein with the domain architecture TM+Band-7+Coiled-coil+Flotillin [**Fig. 3**; **Table S2**; ANH61665.1, Dokdonia], 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 been implicated in resistance to cell-wall targeting antibiotics [96,97]). 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 protein, the YqiJ/YuaF TM protein, and, in some cases, the above-mentioned AAA$^+$-ATPase protein are more widely distributed across bacteria and archaea. These might function along with PspA homologs encoded elsewhere in the genome.

The second major elaboration incorporates the CesT/Tir–PspA gene dyad into a larger 6-7 gene operon (AAN56746.1, Shewanella) [**Fig. 3**; **Table S3**]. A notable feature of these operons is a gene encoding a spermine/spermidine-like synthase domain [98] fused to an N-terminal 7TM transporter-like domain (AAN56744.1, Shewanella; **Fig. 3**; **Table S2**). This operon codes for three additional uncharacterized proteins, including the YjfL family 8TM protein (DUF350 in Pfam), a novel lipoprotein (Ctha_1186 domain), and a β-rich domain that is predicted to localize to the intracellular compartment (AAN56747.1, Shewanella, DUF4178 in Pfam; [**Fig. 3**; **Table S3**]). In some cases, the last gene in this operon codes for a ribbon-helix-helix (RHH) domain transcription factor [**Fig. 3**; **Tables S2, S3**].

The third elaboration that combines the CesT/Tir–PspA gene dyad (AMJ95269.1, Alteromonas) with a polyamine metabolism gene, encodes an ATP-grasp peptide ligase (AMJ95273.1, Alteromonas) related to the glutathionyl spermidine synthetase [**Fig. 3**; **Table S3**]. This association was also recently noticed in a study of AdoMet decarboxylase gene linkages [99]. Additionally, this operon

codes for a potassium channel with intracellular-ligand-sensing TrkA_N and TrkA_C domains, a YjfL family 4TM protein (DUF350 in Pfam; **Table S2**), metal-chelating lipoprotein (DUF1190 in Pfam) that may be distantly related to the eukaryotic SARAF domain involved regulating calcium flux across intracellular membranes, and another uncharacterized protein with potential enzymatic activity (DUF2491 in Pfam). The operons containing spermine/spermidine synthase show two variants when the CesT/Tir–PspA dyad is absent. In the first context carrying DUF4178 (CCP45393.1, Mycobacterium; **Table S3**), the dyad is replaced by a distinct protein occurring as a secreted or lipoprotein version (DUF4247, CCP45395.1, Mycobacterium) and an uncharacterized enzymatic domain with two highly conserved histidines and a glutamate (DUF2617, CCP45394.1, Mycobacterium). In the second, the dyad is replaced by a protein containing a Band-7 domain fused to C-terminal 2TM and SHOCT domains (see below). These operons additionally encode two further polyamine metabolism genes, namely, an AdoMet decarboxylase and a polyamine oxidase.

Another class of associations of PspA, typified by the *Bacillus subtilis ydjFGHL* operon, couples the PspA gene (*ydjF*; ANX09535.1) with multiple other genes that, in firmicutes, includes genes coding for i) a protein with a Band-7 with a C-terminal ZnR (YdjL); and ii) protein with two ZnRs followed by an uncharacterized domain related to YpeB with an NTF2-like fold and a TM segment (YdjG). However, a conserved partner of PspA in these systems is a membrane-associated protein with a so-called "TPM phosphatase domain" (YdjH) [100] [**Fig. 3**; **Table S3**] followed by a C-terminal tail, which might be either a low complexity segment or a long coiled-coil. While this domain has been claimed to be a generic phosphatase based on studies on its plant homolog [100], the evidence supporting this activity is limited. Other studies have implicated the TPM domain in the repair of damaged membrane-associated complexes of the photosystem-II involved in photosynthesis [101,102] and the assembly of the respiratory complex III [103]. Results from our analysis of this domain are more consistent with the latter role. PspA's association with the TPM domain parallels the former's relationship with the chaperone CesT/Tir and Band-7 involved in macromolecular assembly. These observations support an alternative hypothesis that the TPM domain might play a chaperone-like role in the assembly of membrane-linked protein complexes.

Membrane dynamics with Chaperone-like domains

The strong association of PspA with the CesT/Tir type chaperone of two distinct families, the Band-7 and the TPM domains, which are both implicated in the assembly of protein complexes, suggests that these domains might play a role in the assembly of specific membrane-associated complexes along with PspA. These dyads are occasionally part of extended contexts containing the AAA^{+}-ATPase fused to C-terminal coiled-coil and β-propeller domains. Concentrations of polyamines like spermine/spermidine have been previously implicated in membrane stability [104]. The repeated coupling of polyamine metabolism genes in these operons may imply that the PspA-based system for membrane dynamics additionally interfaces with changes in polyamine concentration or aminopropylation of other substrates as recently proposed [99] to alter membrane structure and membrane-associated protein complexes. Based on the repeated coupling with cell-surface lipoprotein, we propose that these systems link extracellular sensing to intracellular membrane dynamics, probably in response to environmental stress conditions. In particular, the flotillin

domain-containing proteins are likely to be recruited to lipid-subdomains with a special composition such as the cardiolipin-rich inner membrane to interface with PspA [97].

Novel PspAA association

The PspA neighborhood analysis identified a new component in the proximity of PspA, which is a protein containing a novel trihelical domain (with absolutely conserved R and D) present in euryarchaeota, thaumarchaeota, actinobacteria, chloroflexi, firmicutes, and a few alpha- and gamma-proteobacteria). This protein occurs in a two-gene cluster with PspA [**Fig. 3**; **Table S3**; *e.g.,* MA_1460; AAM04874.1, Methanosarcina]. This domain mostly occurs by itself but is occasionally fused to an N-terminal PspA in actinobacteria and chloroflexi [**Fig. 3**; **Table S3**; *e.g.,* ACU53894.1, Acidimicrobium]. We call this domain **PspAA** (for **PspA-Associated**; **Table S2**; webapp). Moreover, this PspA–PspAA dyad occurs by itself or in two other contexts: 1) with a two-component system (discussed below; CAB51252.1, Streptomyces) and 2) with another dyad comprising a membrane-associated Metallopeptidase and protein with a novel domain, which we termed **PspAB** (for **PspA-associated protein B**, AAZ55047.1 Tfu_1009 Thermobifida; **Fig. 3**; **Table S2**). This predicted operon occasionally contains a third gene coding for a SHOCT-like (see below) bihelical domain-containing protein in various bacterial and archaeal lineages [**Fig. 3**; **Table S3**; ABW11964.1, Frankia; AKB54760.1, Methanosarcina; AKX93460.1, Moorella].

PspA with PspM (ClgRPspAMN) or Thioredoxin

In cyanobacteria and actinobacteria, PspA may occur in an operon as a dyad with a gene encoding an active Thioredoxin domain-containing protein [17] [**Fig. 3**; **Table S2**]. The actinobacterial PspA found in this dyad is typified by RsmP from *Corynebacterium testudinoris* (AKK09942.1; **Fig. 3**; **Table S3**) that, when phosphorylated, regulates the rod-shaped morphology in *Corynebacterium glutamicum* [105]. This actinobacterial PspA homolog belongs to an RsmP family cluster that predominantly comprises rod-shaped actinobacterial proteins. The PspA homologs in this cluster either occur with **Thioredoxin** (AKK09942.1, Corynebacterium) or with **ClgR-HTH** and **PspM** (CCP45543.1, Mycobacterium; **Table S1**), with an occasional low complexity protein (AOS62694.1, Actinoalloteichus) [**Fig. 3**; **Table S3**]. The Corynebacterial members of the family have paralogs with both versions of PspA contexts ['Paralog' tab of the webapp]. The mutual exclusion of ClgR-PspM and Thioredoxin in the PspA operon suggests that they act as repressors in the case of the HTH and redox regulator in the case of the thioredoxin to control PspA [**Fig. S1**]. The association of ClgR-HTH with PspAM is also confined to this RsmP family, suggesting that these are also determinants of the rod-shaped morphology of the cell. The PspN presence in the immediate operon of ClgR-HTH–PspAM (containing ClgR, PspA, PspM) is limited to a few mycobacteria (CCP45543.1, *M. tuberculosis H37Rv*), which have an N-terminal PspN_N (as defined below) and C-terminal DUF3046. The remaining ClgR-HTH–PspAM operons lack the fused PspN_N–DUF3046 protein and instead contain only the ancestral DUF3046 located three genes downstream (ABK71106.1, *Mycolicibacterium smegmatis*). The duplicated DUF3046 domain forms the intact ClgR-HTH–PspAMN operon only in the *M. tuberculosis* complex [17,18]. The presence of the same family of Thioredoxin with a different family of PspA (typically, two copies) in cyanobacteria suggests that the Thioredoxin homolog is involved in a similar redox activity to control PspA [AFZ14666.1, Crinalium; **Fig. 3**; **Table S3**].

Operons with two-component systems

The association of two-component systems (Receiver/HTH and Histidine Kinase) with the PSP system is widespread. The previously studied *liaIHGFSR* from *B. subtilis* has the following proteins in the operon: LiaI (LiaI-LiaF-TM), LiaH (PspA), LiaG (Toastrack), LiaF (LiaI-LiaF-TM and Toastrack), LiaS (membrane-associated Histidine Kinase), and LiaR (Receiver domain and an HTH) [**Fig. 1, 3**]. This combination is seen mainly in firmicutes and has been studied in the context of lantibiotic resistance [**Fig. 3**; **Table S3**; [21]; ANX06812.1, Bacillus]. In a few Paenibacilli, two genes — an additional PspA and PspC with a C-terminal coiled-coil — have been inserted into a LiaIHGFSR-like operon [**Fig. 3**; **Table S3**; *e.g.,* APB74393.1, part of eight gene context: APB74392.1 to APB74398.1]. In the actinobacterial genus Streptomyces [**Fig. 3**; **Table S3**; CAB51252.1], the dyad of PspA and PspAA occur with a two-component system of the firmicute-type.

Classic Two-component Transcriptional signaling system

Two-component signaling systems utilize either the PspA dominant system in firmicutes and other clades or the PspC dominant system mainly in actinobacteria. As in all two-component systems, the membrane-bound Histidine Kinase communicates a signal based on environmental stress to the receiver domain response regulator (REC+HTH) protein, which regulates the expression of target genes. It is very likely that PspA, LiaI-LiaF-TM, and Toastrack tie into a two-component system to strengthen the membrane in response to the relevant stress signal. In actinobacteria, where the Histidine Kinase is fused to PspC, the signal is presumably sensed by PspC, and Toastrack acts as the internal sensor. Even when these PspC/Toastrack operons with two-component systems lack PspA in their immediate operonic neighborhood, they likely recruit PspA proteins from elsewhere in the genome to bring about the stress response function.

In summary, we have identified novel domain architectures of PspA, including fusions with NlpC and PspAA domains and repeated PspA domains in cyanobacteria [**Fig. 3**]. We have found several PspA paralogs with domain architectures and genomic contexts across bacteria. These characterizations of paralogs are indicative of either gene duplication or horizontal gene transfer ('Paralog' section under 'Phylogeny' tab in the webapp). Further, we have discovered and discussed in some detail our novel findings on genomic contexts housing PspA that are i) extensions or variations of known systems including PspFABC, PspMN, LiaFGRS, and ii) novel themes including AAA^{+}-ATPase (with PspA/Snf7), PspAA, mutual exclusion with PspM and Thioredoxin, chaperones, and two-component systems [**Fig. 3**; PspA/Snf7 selection in all tabs of the webapp]. Following the detailed analyses of PspA that uncovered novel insights into the many variations in its genomic contexts, we proceeded to analyze the domain architectures of Psp partner proteins to obtain a more holistic understanding of the PSP stress response system.

PspA-free variations of domain architectures and gene-neighborhood

Conducting searches with multiple proteins and their domains also shed light on the existence of several domain architectures and genomic contexts involving Psp components that do not carry the central protein PspA. Hence, we next investigated these instances to determine their possible

relevance to the extended PSP stress response systems. Below we describe some of our significant findings related to the PspC and Toastrack domains and their homologs, and then we outline a range of novel PspA-free genomic contexts.

PspC domain architectures and gene-neighborhood

PspC by itself is present in most bacteria and the archaeal clades of euryarchaeota and Asgardarchaeota. Some orthologs of PspC are fused to an N-terminal ZnR domain sporadically in various clades [**Table S4**; ABC83427.1, Anaeromyxobacter]. Almost all other PspC homologs either show fusions to, or occur in predicted operons coding for, one or more TM domains such as the LiaI-LiaF-TM and PspB [**Fig. 4**; **Table S4**]. Additionally, PspC is also fused to diverse signaling domains such as i) the Histidine Kinase from two-component systems (see above), ii) a novel signaling domain that we term the HTH-associated α-helical signaling domain (**HAAS**, overlaps partly with the Pfam model for the domain of unknown function, DUF1700; **Table S2**), and iii) the Toastrack domain (see below for the latter two domains; **Table S1**). Moreover, actinobacteria contain a Lia-like system without a PspA gene. The histidine kinase in these systems is distinct from that seen in the Lia-like systems with PspA. This histidine kinase is fused to N-terminal PspC and LiaF-LiaI-TM domains (*e.g.,* AIJ13865.1, Streptomyces) and is accompanied by a Receiver domain fused to a DNA-binding HTH domain [**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, Streptomyces; **Fig. 5**; **Table S4**]. These associations strongly imply that PspC is a sensor domain that likely senses a membrane-proximal signal and communicates via the associated domains.

Contextual associations of the Toastrack Domain

The **Toastrack** domain repeatedly emerges as a partner to the PSP system. As noted above, a gene coding for a Toastrack protein widely co-occurs with PspA in several conserved gene neighborhoods [**Fig. 4**; **Table S4**; see *Methods*] and is also fused to PspC when it occurs independently of PspA. We found that the Toastrack domains tend to be C-terminal to various TM domains such as PspC, LiaI-LiaF-TM, other multi-TM domains unique to these systems, or single TM anchors [**Fig. 4**]. In cyanobacteria, we find variable multidomain proteins with an N-terminal TM anchor followed by a region containing the Toastrack domain flanked by immunoglobulin (Ig) and one or more catalytic domain such as a Fringe-like glycosyltransferase or a caspase-like thiol peptidase [**Fig. 4**; **Table S4**; AFY83227.1, Oscillatoria; **Table S2**]. These configurations strongly suggest that the Toastrack is an intracellular domain. Further, in several architectures, the N-terminal TM regions fused to the Toastrack domain are replaced by at least two variants of the bihelical SHOCT (e.g., *Bacillus subtilis* ***yvlB*** (CAB15517.1, Bacillus) [**Fig. 4, 5**; **Table S4**]. We call these variants **SHOCT-like** domains to distinguish them from the classical SHOCT domain, as these include a domain partly detected by the Pfam **DUF1707** [106] model and another that has not been detected by any published profile. The SHOCT and related domains are fused to disparate domains and are typically found at the N- or C-termini of proteins. Based on this observation and displacement of the TM regions by the SHOCT-like domains, we predict that this domain plays a role in anchoring disparate domains, including Toastrack, to the inner leaf of the membrane. Moreover, the remaining Toastrack domains found in multidomain architectures are fused to N-terminal HTH domains. The DNA-binding domain

LytTR may also be encoded by operons with Toastrack fused to LiaI-LiaF-TM in firmicutes Clostridium (CAL82154.1) and Lactobacillus (AGK93623.1) [**Fig. 5**; **Tables S2, S4**].

Furthermore, the YvlB/YthC Toastrack protein (*e.g.,* CAB15517.1, Bacillus), which occurs in an operon with a PspC homolog (ythA/ythB/YvlC; CAB15516.1), has been implicated in the activation of the membrane-associated LiaFSR operon [107]. The ythA/ythB/YvlC family of PspC has been shown to contribute to protection against membrane permeabilization [64,108]. We also identified variants of the classic Lia operon with Toastrack and the two-component system that do not contain PspA, typified by *vraT* of *Staphylococcus aureus* [ABD31150.1; **Fig. 5**; **Table S4**]. These systems carry LiaI-LiaF-TM and Toastrack with the two-component system (*vraSR*) and are involved in methicillin and cell-wall-targeting antibiotic resistance [3,109–111]. A similar organization of LiaI-LiaF-TM and Toastrack containing protein with a two-component system is found in bacteroidetes, a few acidobacteria, beta-, gammaproteobacteria (ABD83157.1), and Ignavibacteriae (AFH48155.1) [**Table S4**].

We also recovered several other conserved gene neighborhoods centered on Toastrack genes that are in contexts that are likely to define novel functional analogs of the Psp system with potential roles in membrane-linked stress response: The first of these found across diverse bacterial lineages contains a core of four genes coding for i) a Sigma factor, ii) a receptor-like single TM protein with an intracellular anti-sigma-factor zinc finger (zf-HC2, PF13490 in Pfam) and extracellular HEAT repeats, iii) one or two membrane-anchored Toastrack-containing proteins (AFK03672.1 Emticicia; **Table S4**), and iv) a previously uncharacterized protein with hits to the Pfam model DUF2089. We found that this Pfam model **DUF2089** can be divided into an N-terminal ZnR, central HTH, and C-terminal SHOCT-like domains (ADE70705.1, Bacillus) [**Fig. 4**; **Tables S2, S4**]. In a few of these operons, the membrane anchor of the Toastrack domain is a LiaI-LiaF-TM domain [Fig. 5; Table S4]. Variants of this system include additional genes coding for a protein with a LiaI-LiaF-TM domain fused to an N-terminal B-box domain (*e.g.,* AEU34960.1, Granulicella) or a PspC protein (*e.g.,* OGF50123.1 from Candidatus Firestonebacteria) [**Fig. 5**; **Table S4**]. We propose that this three-gene system functions similarly to the classical Lia operon in transducing membrane-associated signals to a transcriptional output affecting a wide range of genes via the sigma factor.

Similarly, an operon observed predominantly in various proteobacteria and bacteroidetes couples a protein with a membrane-anchored Toastrack domain (typified by AAM36414.1, Xanthomonas) with genes coding for an ABC-ATPase, a permease subunit, and a **GNTR-HTH** transcription factor with distinct C-terminal α-helical domain and another [**Fig. 5**; **Tables S2, S4**]. These operons also code for a previously uncharacterized protein matching the Pfam DUF2884 model. We show that these proteins are membrane-associated lipoproteins (*e.g.,* AJI31452.1, Bacillus), which might function as an extracellular solute-binding partner for the ABC-ATPase and permease components. A comparable operon found in actinobacteria replaces the **GNTR-HTH** transcription factor with a Ribbon-helix-helix (**RHH**) domain protein. In some actinobacteria, the Toastrack domain encoded by the operon is fused to a SHOCT-like domain and is encoded adjacent to genes specifying a two-component system (CCP43715.1, Mycobacterium) or a transport operon (CAB88834.1, Streptomyces) [**Fig. 5**; **Table S4**]. These operons with the Toastrack domains are likely to couple

transcriptional regulation to the sensing of membrane-proximal signal and transport [**Fig. 5**; **Table S4**]. The GNTR-HTH and RHH operons in these systems are likely to function as transcriptional regulators analogous to PspF and ClgR transcription factors from classical Psp systems.

These observations support our proposal that the Toastrack domains are likely localized proximal to the membrane. The two extended β-sheets of the Toastrack domain (formed by the β-helical repeats) present two expansive surfaces that are amenable to protein-protein interactions. Hence, we suggest that this domain provides a platform for the assembly of sub-membrane signaling complexes. The docking of proteins to the Toastrack domain could be further transduced to activate transcription via fused or associated HTH or LytTR domains, associated two-component systems, or the HAAS-HTH system (see below).

The HAAS-PadR-like wHTH systems

We discovered a novel signaling system that frequently co-occurs with PspC and Toastrack domains as part of gene neighborhoods that are mostly independent of PspA [**Fig. 4, 5**]. This system consists of two components: the **HAAS** domain and the **PadR-like wHTH** [**Table S2**]. The HAAS domain is α-helical (partly detected by Pfam models DUF1700, DUF1129; Pfam clan: Yip1). An HHpred profile-profile search further unifies this model with the Pfam profile DUF1048 (PDB: 2O3L). We, therefore, unify these models together as the HAAS superfamily [see HAAS multiple sequence alignment as part of the 'MSA' tab of the webapp]. The core HAAS fold has three consecutive α-helices, with the third helix displaying a peculiar kink in the middle corresponding to a conserved GxP motif, which is 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-occurrence suggests that HAAS transduces a signal from the domain fused to it to this partner transcription factor with the PadR-like wHTH domain.

When contextually linked to PspC, the HAAS domain shows three broad themes: 1) as part of a multidomain TM protein with additional PspC, LiaI-LiaF-TM, and Toastrack domains; 2) fused directly to a Toastrack domain in a gene neighborhood that also codes for a PspC domain protein occurring either by itself or fused to other TM domains; 3) as part of a TM protein fused to conserved multi-TM domains other than the LiaI-LiaF-TM or PspC (*e.g.,* VanZ) [**Fig. 4, 5**; **Table S4**]. These gene neighborhoods code for standalone PspC genes. In addition to these associations, the HAAS domains occur in contexts independent of PspC but are typically coupled to the N-terminus of other multi-TM or intracellular sensory domains. We found conserved fusions to several distinct multi-TM domains, such as: 1) The FtsX-like TM domains with extracellular solute-binding MacB/PCD domains [**Fig. 4, 5**; **Table S4**; ACO32024.1, Acidobacterium]; 2) FtsW/RodA/SpoVE family of TM domains [**Fig. 4, 5**; **Table S4**; CAC98500.1, Listeria] [112]; 3) various uncharacterized 6TM, 4TM and 2TM domains. In addition to Toastrack, the HAAS domain may fuse to other intracellular domains such as the Pentapeptide repeat domains [AOH56696.1, Bacillus; **Fig. 5**; **Table S4**] (mostly in firmicutes). While these operons do not contain PspC, the organisms have Psp components elsewhere in the genome. These proteins might be recruited to the stress response system as suggested by the effects of PadR-like wHTH deletion studies in the Listeria Lia systems [112].

A key observation is that the HAAS–PadR-like-wHTH dyad takes the place of the histidine kinase-receiver domain transcription factor dyad in bacteria that do not contain the typical Lia operon (carrying the LiaRS-like two-component system). Hence, we posit that the HAAS-PadR dyad constitutes an alternative to the classical two-component systems. Here, the HAAS domain could 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 associated with the conserved GxP motif; this change might lead to the release of the transcription factor to mediate a downstream transcriptional response.

In summary, we have discovered several contextual connections that expand the network of Psp partner domains. These suggest that the broader Psp-associated network might feature a repertoire of proteins beyond those previously studied in Psp signaling.

A unified view of PSP partners and evolution

As described in the sections above, we have identified various genomic configurations of the PSP systems and novel partner domains that likely function alongside or independent of PspA. We have also determined key neighbors of PspC and Toastrack, the most wide-spread PspA partners, as well as the domains that frequently co-occur with our initial set of proteins of interest. As our final question, we asked how to reconcile all these pieces of the Psp puzzle.

The proximity network of PspA and cognate partner domains

To generate a global picture, we first built a network representation of the domain architectures of the extended PSP system reported above [**Fig. 6A**]. The domains are the nodes of this network; pairs of domains are connected if they co-occur in a single protein. The network summarizes the connections between PspA, Psp cognate partner domains (such as PspC and Toastrack), and other novel domain connections such as HAAS and SHOCT-bihelical/SHOCT-like proteins.

The phyletic spread of Psp members and their prevalent partner domains

Next, to bring together the evolutionary patterns of all Psp members and their key partners, we created a single heatmap that displays their phyletic spreads together with their prevalence across lineages [**Fig. 6B**]. This view recapitulates the findings that (i) only PspA, Snf7, and Toastrack are present in eukaryotes; (ii) while PspC, PspA/Snf7 are present in most archaeal lineages, our analysis identifies occasional transfers of Toastrack, LiaI-LiaF-TM, and SHOCT-bihelical to euryarchaeota; (iii) in bacteria, domains such as Toastrack, PspC, PspA, and LiaI-LiaF-TM co-migrate.

The prevalence of most common neighbors of Psp members

Finally, we summarized the overlap of several sets of domain architectures concurrently. Using an 'upset plot' visualization, we quantified the relative occurrence of domain architectures from Psp members and their most common neighbors [**Fig. 6C**, blue histogram] as well as combinations [**Fig. 6C**; dots and connections] and their frequencies [**Fig. 6C**, red histogram] of co-occurrences of these

domain architectures in specific genomic contexts. For example, we observe that singleton PspA/Snf7, PspC, and Toastrack domains with uncharacterized neighborhoods are most prevalent. These are closely followed by HAAS–PadR-like-wHTH dyads with/without Toastrack, the Toastrack/LiaI-LiaF-TM contexts with two-component systems, and PspABC operons [**Fig. 6C**]. We note that Snf7 appearing alone most frequently is an artifact of the genomic context analysis being restricted (and relevant) to bacteria and archaea (see webapp).

In summary, these visualizations together provide a consolidated picture of all the aspects of the much-expanded Psp universe as ascertained by our detailed molecular evolution and phylogenetic analysis [**Fig. 6**; webapp].

Conclusions

Recent work with the PSP stress response system in bacteria suggests that even though its function appears to be maintained across phyla, the linked accessory Psp proteins, the regulatory mechanisms, and the membrane stress response dynamics of the system vary widely among bacterial species [17,20,25,36]. PspA has also been shown to i) help maintain the cell shape in Corynebacterium [105], and be involved in ii) cell division in archaea [34], and iii) thylakoid biogenesis and vesicular budding in eukaryotes [113]. These variations likely reflect different ways in which a common peripheral membrane protein is utilized in lineage-specific envelope dynamics and responses to the environmental stresses uniquely experienced by different lineages.

In this study, we report the results of the first comprehensive analysis of the evolution of the PSP system and all its partner proteins, focusing on the phyletic distribution, sequence/structural features of PspA, its known and unknown partners, and their genomic neighborhoods across the tree of life. We used these analyses to explore and identify novel components of the Psp stress response system. We first established that PspA/Snf7 is universal and that the ancestry of the PspA/Snf7 superfamily traces back to the last universal common ancestor (LUCA), in agreement with a recent study [10]. A corollary to this finding is that, despite the different types of membranes (*e.g.,* ether-linked vs. ester-linked) in the archaeal, bacterial, and eukaryotic lineages, the LUCA already 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 inferences based on the signal-recognition particle GTPases [114] and ATP-synthases implying the presence of a transmembrane secretory and ion-transport apparatus in the LUCA [115,116]. These inferences suggest that this ancient machinery was robust to the subsequent changes in membrane composition that occurred in different lineages.

In addition to characterizing the global phyletic distribution of PspA and its known partners, we discovered novel themes (domain architectures and genomic contexts) in PSP systems that are either widely distributed (*e.g.,* PspA/Snf7, PspC, and Toastrack) or restricted to specific lineages (*e.g.,* HAAS and SHOCT-bihelical were often proximal to Toastrack, mostly in firmicutes and actinobacteria, respectively). Finally, in addition to novel extensions and variations of known Psp

operons (with PspBC, LiaI-LiaF-TM/Toastrack, and two-component systems), we discovered several novel components such as new transcriptional signaling systems and chaperone systems.

Several of these new findings are of particular importance: 1) Based on its occurrence with membrane partners (such as PspA, LiaI-LiaF-TM, PspC, HAAS, SHOCT-like domains) and transcription regulators/two-component systems (such as LiaRS, PadR-like-wHTH), we propose that Toastrack is an intracellular domain likely proximal to the membrane that provides a platform for the assembly of sub-membrane signaling complexes. 2) The finding that the PspA homologs have repeatedly associated with distinct AAA^{+}-chaperones as well as a non-ATPase chaperone or predicted chaperone-like proteins (*e.g.,* CesT/Tir, Band-7, and TPM) suggests that these associations are involved in the assembly of sub- and trans-membrane complexes presumably in response to specific envelope stresses. These observations suggest a more general role for ESCRT-like complexes in these bacterial systems. 3) The association of PspA with various transporters and polyamine metabolism systems suggests that the regulation of membrane structure by PspA is associated with changes in the concentration of different solutes that may affect membrane stability. 4) The finding of several new versions of the SHOCT-like fold suggests that the SHOCT-like domains mediate membrane localization of disparate catalytic and non-catalytic activities, which in turn may interface with the PSP system. 5) A diverse array of sub-membrane (*e.g.,* Toastrack) and surface domains (*e.g.,* PspC, LiaI-LiaF-TM) co-occur with two-component or HAAS–PadR-like-wHTH systems, which suggests alternatives to conventional two-component signaling that interface with the PSP system in signaling membrane stress.

All the findings (data, results, and visual summaries) from this work are found in an interactive and queryable web application available at https://jravilab.shinyapps.io/psp-evolution.

Methods

We used a computational evolutionary approach for the molecular characterization and neighborhood analyses of the Psp stress-response system and its partners.

Query and subject selection

All known Psp members — PspA (with eight representatives across Proteobacteria, Actinobacteria, two copies from Firmicutes, cyanobacteria, and archaea); PspM (Rv2743c) and PspN (Rv2742c) (from *M. tuberculosis*); PspB and PspC (from *E. coli*); LiaI, LiaG, and LiaF (from *B. subtilis*) — were queried against all sequenced genomes across the tree of life. Homology searches were run against custom databases of ~6500 completed representative/reference genomes with taxonomic lineages or the NCBI NR database [117,118] [**Table S1**]. All Psp homologs are listed in the webapp, with representatives in **Tables S3, S4**. The phyletic order (sequence) was obtained from NCBI taxonomy, PATRIC [117–122].

Identification and characterization of protein homologs

To ensure the identification of a comprehensive set of homologs (close and remote) for each queried protein, we performed iterative searches using PSIBLAST [123] with sequences of both full-length

proteins and the corresponding constituent domains. For each protein, searches were conducted using homologous copies from multiple species as starting points. Search results were aggregated, and the numbers of homologs per species and genomes carrying each of the query proteins were recorded [webapp]. These proteins were clustered into orthologous families using the similarity-based clustering program BLASTCLUST [124]. HHPred, SignalP, TMHMM, Phobius, JPred, Pfam, and custom profile databases [125–132] were used to identify signal peptides, TM regions, known domains, and the secondary protein structures in every genome. Homolog information, including domain architectures, can be accessed in the webapp ('Data,' 'Domain architectures' tabs).

Neighborhood search

Bacterial gene neighborhoods — ±7 genes flanking each protein homolog — were retrieved using custom scripts from GenBank [117,118]. Gene orientation, domains, and secondary structures of the neighboring proteins were characterized using the same methods applied to query homologs above. Genomic contexts can be accessed in the webapp ('Genomic contexts' tab).

Phylogenetic analysis

Multiple sequence alignment (MSA) of the identified homologs was performed using Kalign [133] and MUSCLE [134]. The phylogenetic tree was constructed using FastTree 2.1 with default parameters [135]. MSA and phylogenetic trees can be accessed in the webapp ('Phylogeny' tab).

Network reconstruction

The Psp proximal neighborhood network was constructed based on the domain architectures and genomic contexts of PspA and its cognate partner proteins [**Tables S1, S2**]. The nodes represented the domains, and edges indicated a shared neighborhood (domains of the same protein or neighboring proteins). Proximity networks can be accessed in the webapp ('Domain architectures' tab).

Web application

The interactive and queryable web application was built using R Shiny [136,137]. All data analyses and visualizations were carried out using R/RStudio [138,139] and R-packages [140–153]. All the data and results from our study are available on the web application (https://jravilab.shinyapps.io/psp-evolution).

Figures

Figure 1. The phyletic spread of known Psp members across all major lineages.

A. The three known PSP systems in *E. coli* (pspF||pspABC), *M. tuberculosis* (clgRpspAMN), and *B. subtilis* (lialHGFSR) are shown. Boxes indicate genes/proteins, and colors indicate the nature of the protein (black, PspA homolog; teal, transcription factor/regulator; warmer shades of orange/yellow, transmembrane protein). Thin arrows denote the direction of transcription, and block arrows denote promoters. **B. Domain architectures of the known Psp operons** in *E. coli*, *M. tuberculosis*, and *B. subtilis*. Domains are denoted by rectangular segments inside the block arrow representing a protein. 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 for the homologs of 'known' 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 kingdoms, bacteria, archaea, and eukaryota. The outer ring depicts the distribution of the homologs among key lineages (phyla). Interactive sunburst plots for each of the Psp proteins are available in the webapp.

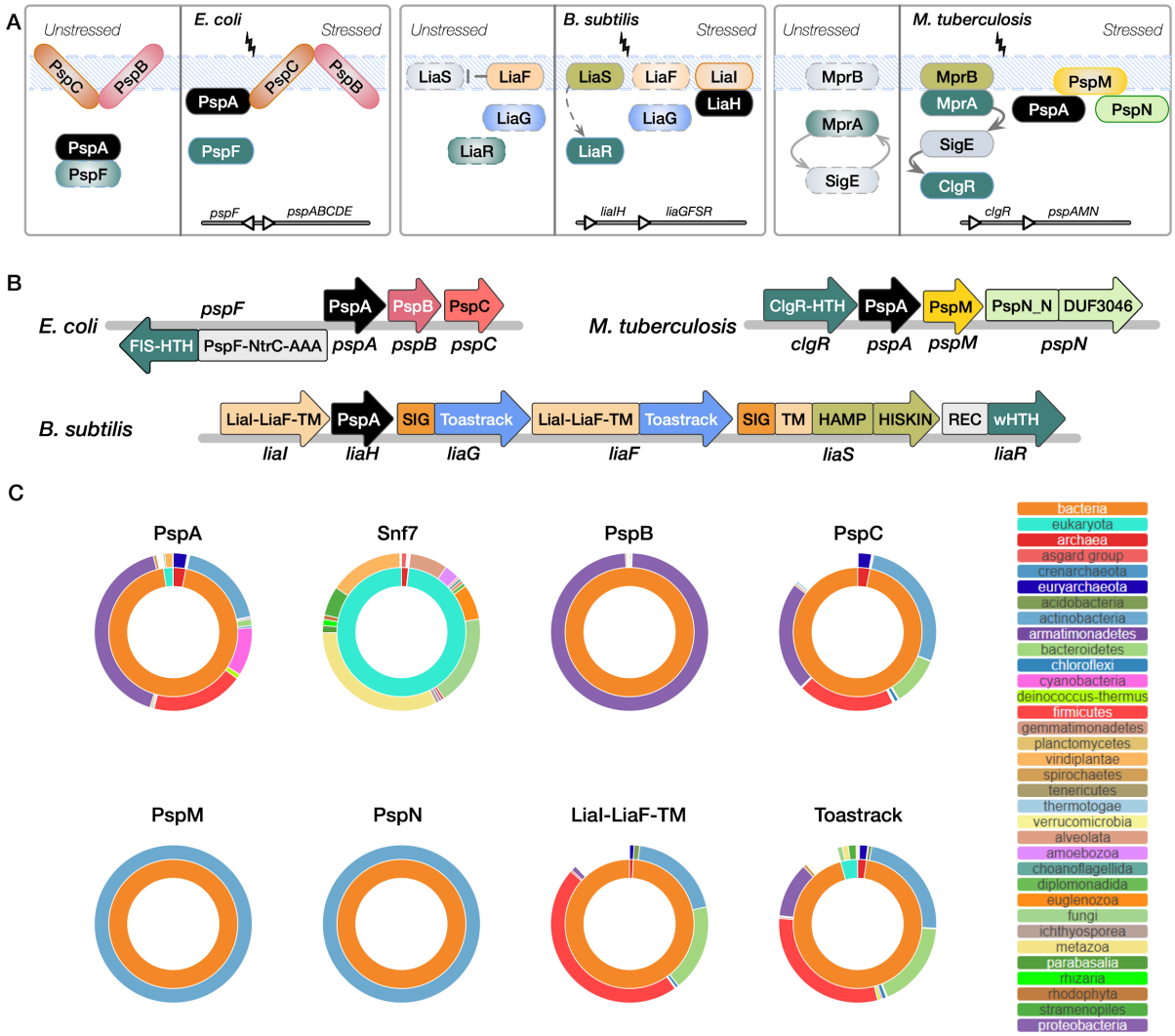


Figure 2. PspA/Snf7 homologs across the tree of life.

A. Phylogenetic tree of PspA homologs across the tree of life. The phylogenetic tree was constructed based on a multiple sequence alignment performed using representative PspA homologs across all the major kingdoms/phyla (see *Methods*; webapp). The key lineages are labeled next to distinct clusters of similar PspA proteins. 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. The outgroup at the bottom of the tree includes Snf7 homologs. **B. Phyletic spreads of PspA/Snf7 domain architectures across the tree of life.** The phyletic spread of the various PspA/Snf7 domain architectures is shown along with their relative frequencies as a stacked barplot. Further details of the domain architectures of all PspA/Snf7 homologs and their phyletic spreads are shown in the webapp, with representative ones shown in **Table S3**.

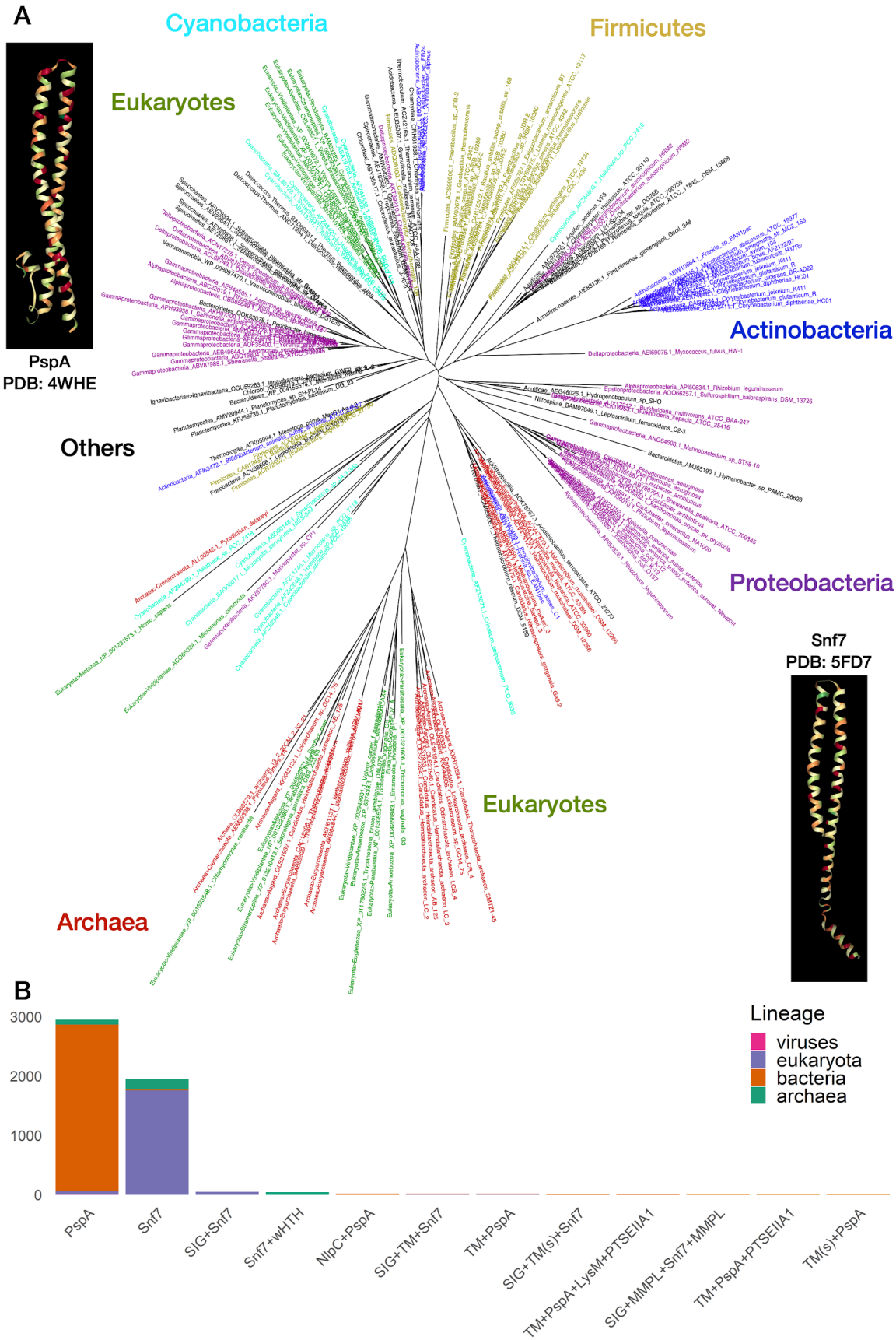

A
Cyanobacteria
Firmicutes
Eukaryotes
PspA
PDB: 4WHE
Actinobacteria
Others
Proteobacteria
Snf7
PDB: 5FD7
Eukaryotes
Archaea
B
3000
2000
1000
0
Lineage
viruses
eukaryota
bacteria
archaea
PspA
Snf7
SIG+Snf7
Snf7+wHTH
NlpC+PspA
SIG+TM+Snf7
TM+PspA
SIG+TM(s)+Snf7
TM+PspA+LysM+PTSEIIA1
SIG+MMPL+Snf7+MMPL
TM+PspA+PTSEIIA1
TM(s)+PspA

Figure 3. PspA domain architectures and genomic contexts

The first row contains domain architectures of the most prevalent homologs of PspA (indicated in black throughout). The remaining rows show the predominant genomic contexts of PspA homologs across multiple bacterial and archaeal lineages. Representative neighborhoods, along with their 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) CIgR, PspM/PspN, Thioredoxin [**Fig. S1**]; (d) chaperones such as Band-7, 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. Representation of the predominant contexts that were identified by neighborhood searches (±7 genes flanking each homolog; see *Methods*). 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 different colors and by the same coloring scheme across the figures in this study. 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 in **Table S3**, and the full list of PspA homologs, their domain architectures, genomic contexts, and lineages are shown in the webapp (under the 'Data,' 'Domain architectures,' and 'Genomic contexts' tabs).

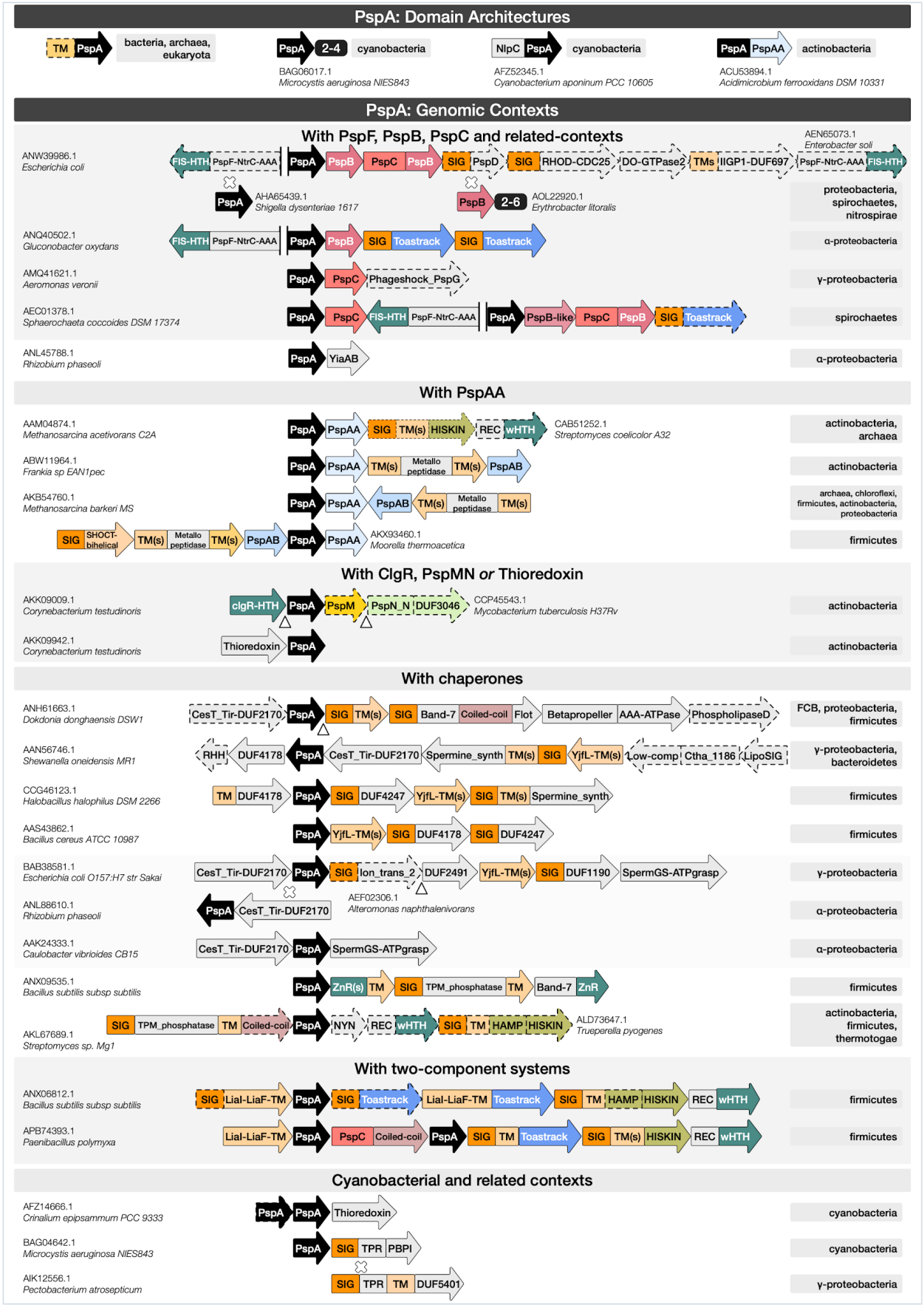

PspA: Domain Architectures
TM PspA
bacteria, archaea, eukaryota
PspA 2-4
cyanobacteria
BAG06017.1
Microcystis aeruginosa NIES843
NlpC PspA
cyanobacteria
AFZ52345.1
Cyanobacterium aponinum PCC 10605
PspA PspAA
actinobacteria
ACU53894.1
Acidimicrobium ferrooxidans DSM 10331
PspA: Genomic Contexts
With 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 IIIGP1-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
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 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

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 known (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. The phyletic spread of the various domain architectures is shown along with their relative frequencies as a stacked barplot. Further details of the domain architectures of all PspA partner domain homologs and their phyletic spreads are shown in the webapp, with representative ones shown in **Table S4**.

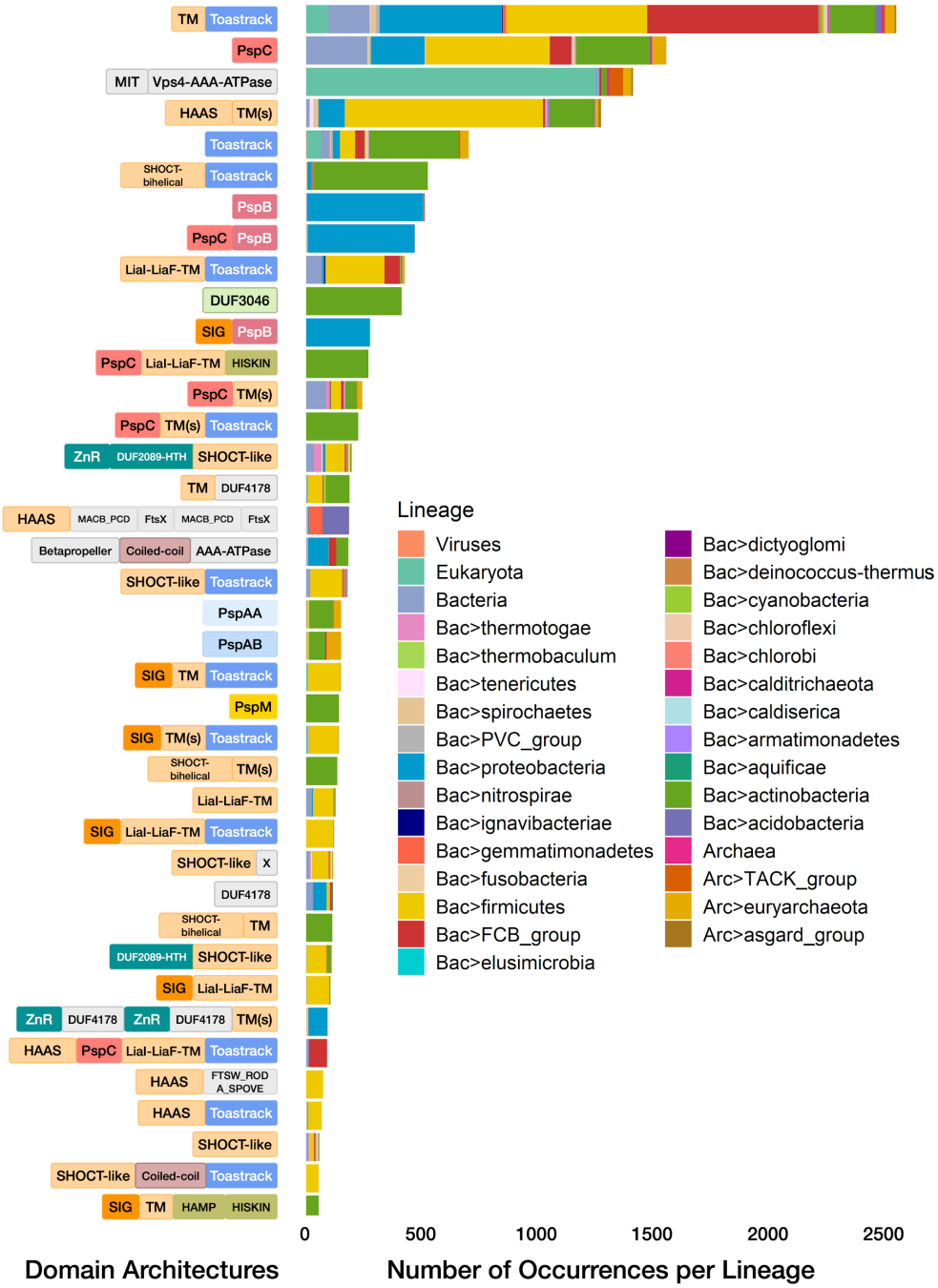


Figure 5. The genomic contexts housing all Psp cognate partner domain homologs.

The genomic contexts 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-bihelical (in shades of orange), 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 webapp, and representatives indicated in the figure are shown in **Table S4**.

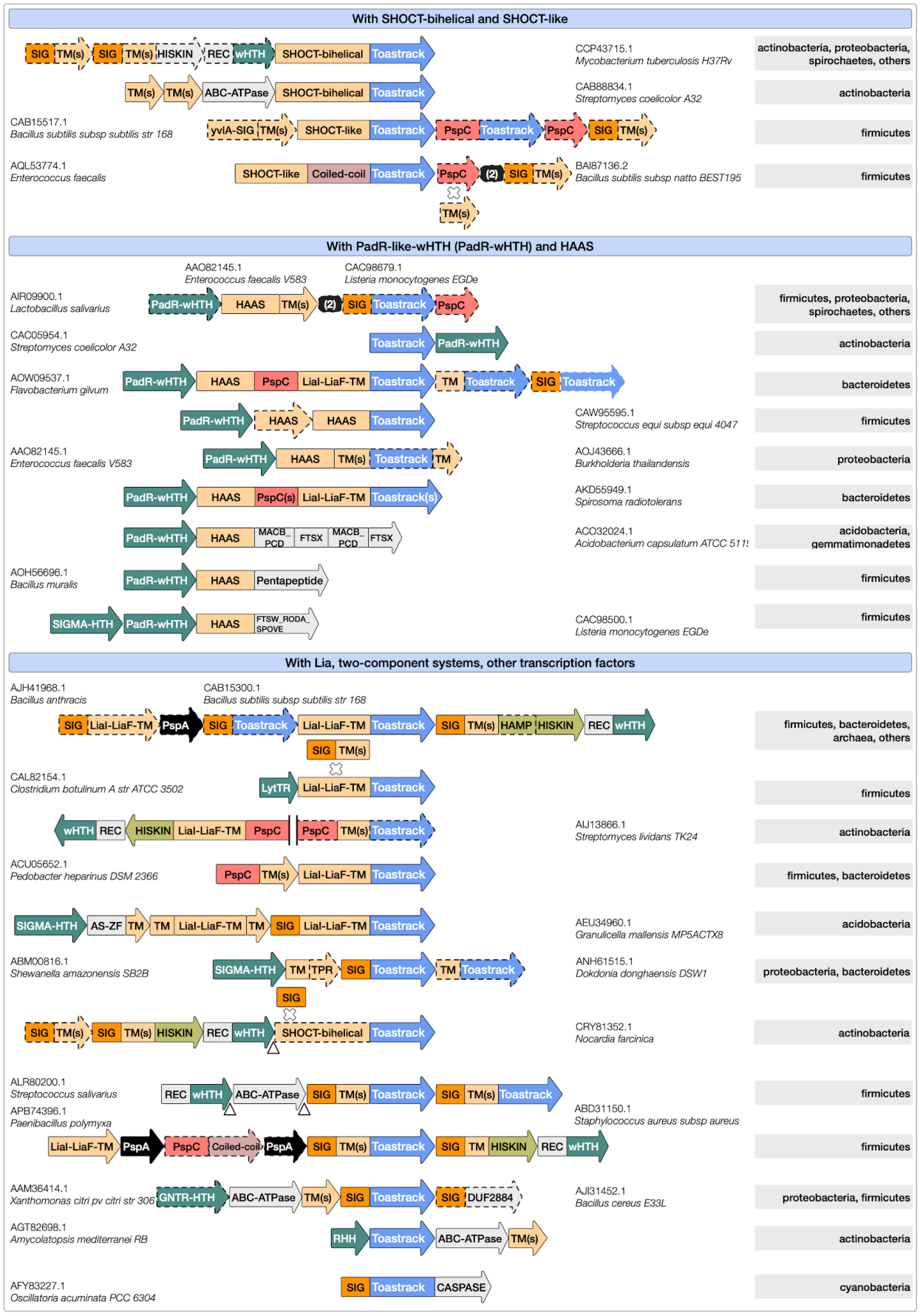
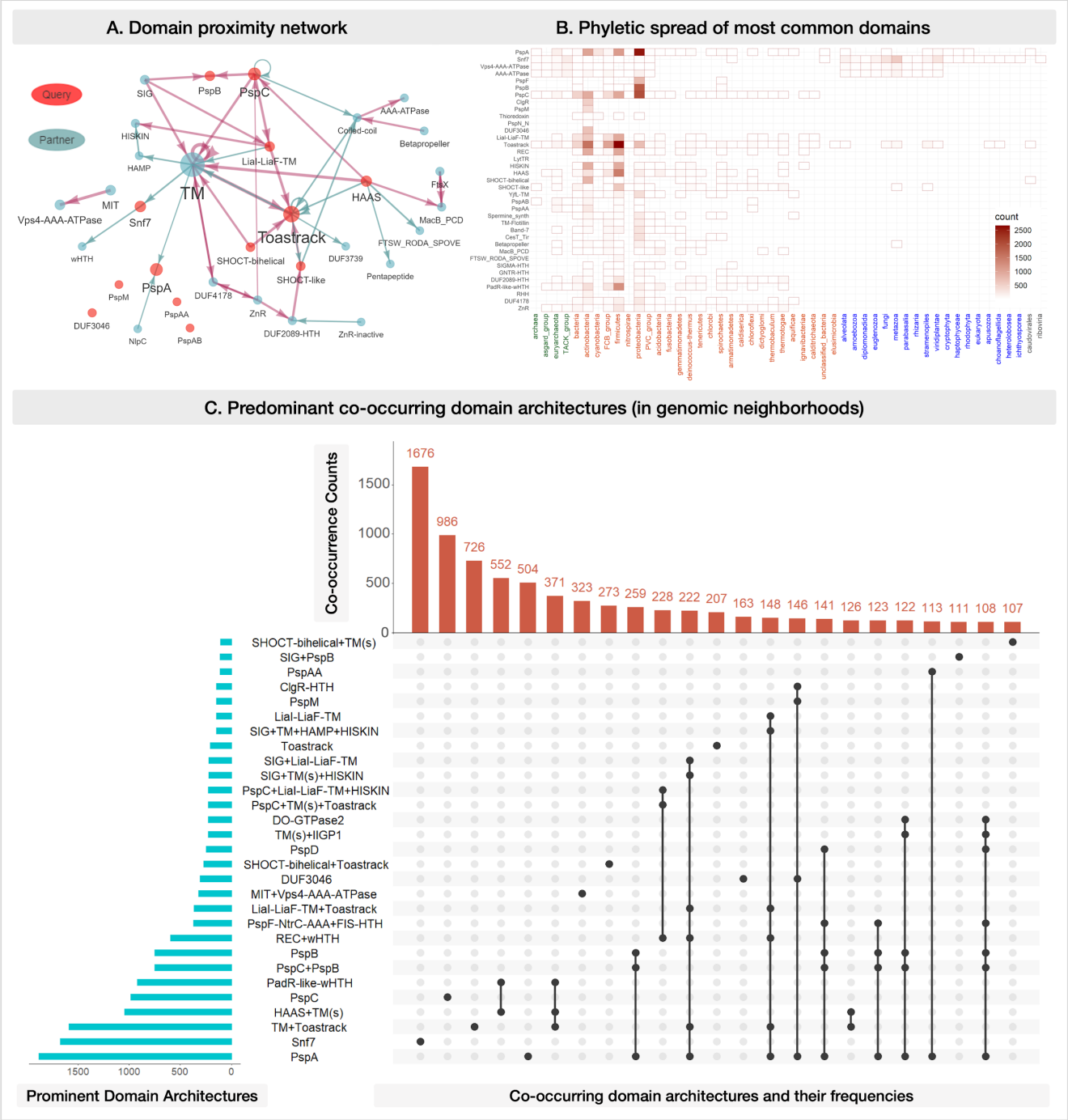

With SHOCT-bihelical and SHOCT-like
SIG TM(s) SIG TM(s) HISKIN REC wHTH SHOCT-bihelical Toastrack
CCP43715.1
Mycobacterium tuberculosis H37Rv
actinobacteria, proteobacteria, spirochaetes, others
TM(s) TM(s) ABC-ATPase SHOCT-bihelical Toastrack
CAB88834.1
Streptomyces coelicolor A32
actinobacteria
CAB15517.1
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
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

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 proportionate 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 indicated in red and grey. *Note*: A few connections may be absent in the network here 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, are available in the webapp. **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). *Rows*: Psp members and their most frequent partners are queried against all sequenced and completed genomes across the three major kingdoms of life. *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 webapp (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) underlying 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 webapp.


A. Domain proximity network
Query
Partner
SIG
PspB
PspC
AAA-ATPase
Coiled-coil
Betapropeller
HISKIN
HAMP
Lial-LiaF-TM
TM
HAAS
FtsX
MIT
Vps4-AAA-ATPase
Snf7
MacB_PCD
Toastrack
FTSW_RODA_SPOVE
wHTH
SHOCT-bihelical
DUF3739
SHOCT-like
Pentapeptide
PspA
PspM
DUF4178
PspAA
ZnR
DUF3046
NlpC
PspAB
DUF2089-HTH
ZnR-inactive
B. Phyletic spread of most common domains
count
2500
2000
1500
1000
500
C. Predominant co-occurring domain architectures (in genomic neighborhoods)
Co-occurrence Counts
1676
986
726
552
504
371
323
273
259
228
222
207
163
148
146
141
126
123
122
113
111
108
107
SHOCT-bihelical+TM(s)
SIG+PspB
PspAA
ClgR-HTH
PspM
Lial-LiaF-TM
SIG+TM+HAMP+HISKIN
Toastrack
SIG+Lial-LiaF-TM
SIG+TM(s)+HISKIN
PspC+Lial-LiaF-TM+HISKIN
PspC+TM(s)+Toastrack
DO-GTPase2
TM(s)+IIGP1
PspD
SHOCT-bihelical+Toastrack
DUF3046
MIT+Vps4-AAA-ATPase
Lial-LiaF-TM+Toastrack
PspF-NtrC-AAA+FIS-HTH
REC+wHTH
PspB
PspC+PspB
PadR-like-wHTH
PspC
HAAS+TM(s)
TM+Toastrack
Snf7
PspA
1500
1000
500
0
Prominent Domain Architectures
Co-occurring domain architectures and their frequencies

Declarations

Acknowledgments

We are incredibly grateful to Arjun Krishnan and Krishnan Raghunathan for detailed comments on the manuscript and web-application, guidance, and support. 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.

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 (MSU) College of Veterinary Medicine Endowed Research Funds, and MSU 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. 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.shinyapps.io/psp-evolution/. Text, figures, and the webapp are licensed under Creative Commons Attribution CC BY 4.0.

References

1. van den Brink-van der Laan E, Killian JA, de Kruijff B. Nonbilayer lipids affect peripheral and integral membrane proteins via changes in the lateral pressure profile. Biochim. Biophys. Acta 2004; 1666:275–288
2. Kobayashi R, Suzuki T, Yoshida M. Escherichia coli phage-shock protein A (PspA) binds to membrane phospholipids and repairs proton leakage of the damaged membranes. Mol. Microbiol. 2007; 66:100–109
3. Jordan S, Hutchings MI, Mascher T. Cell envelope stress response in Gram-positive bacteria. FEMS Microbiol. Rev. 2008; 32:107–146
4. Raivio TL. Envelope stress responses and Gram-negative bacterial pathogenesis. Mol. Microbiol. 2005; 56:1119–1128
5. Rowley G, Spector M, Kormanec J, et al. Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens. Nat. Rev. Microbiol. 2006; 4:383–394
6. Omardien S, Drijfhout JW, van Veen H, et al. Synthetic antimicrobial peptides delocalize membrane bound proteins thereby inducing a cell envelope stress response. Biochim Biophys Acta Biomembr 2018; 1860:2416–2427
7. Wenzel M, Kohl B, Münch D, et al. Proteomic response of Bacillus subtilis to lantibiotics reflects differences in interaction with the cytoplasmic membrane. Antimicrob. Agents Chemother. 2012; 56:5749–5757
8. Guest RL, Wang J, Wong JL, et al. A Bacterial Stress Response Regulates Respiratory Protein Complexes To Control Envelope Stress Adaptation. J. Bacteriol. 2017; 199:
9. Gottesman S. Stress Reduction, Bacterial Style. J. Bacteriol. 2017; 199:
10. Liu J, Tassinari M, Souza DP, et al. Bacterial Vipp1 and PspA are members of the ancient ESCRT-III membrane-remodelling superfamily. bioRxiv 2020; 2020.08.13.249979
11. Lee I-H, Kai H, Carlson L-A, et al. Negative membrane curvature catalyzes nucleation of endosomal sorting complex required for transport (ESCRT)-III assembly. Proc. Natl. Acad. Sci. U.S.A. 2015; 112:15892–15897
12. Mehner D, Osadnik H, Lünsdorf H, et al. The Tat system for membrane translocation of folded proteins recruits the membrane-stabilizing Psp machinery in Escherichia coli. J. Biol. Chem. 2012; 287:27834–27842
13. McDonald C, Jovanovic G, Ces O, et al. Membrane Stored Curvature Elastic Stress Modulates Recruitment of Maintenance Proteins PspA and Vipp1. mBio 2015; 6:e01188-01115
14. McDonald C, Jovanovic G, Wallace BA, et al. Structure and function of PspA and Vipp1 N-terminal peptides: Insights into the membrane stress sensing and mitigation. Biochim Biophys Acta Biomembr 2017; 1859:28–39
15. Kleerebezem M, Crielaard W, Tommassen J. Involvement of stress protein PspA (phage shock protein A) of Escherichia coli in maintenance of the protonmotive force under stress conditions. EMBO J. 1996; 15:162–171
16. Heidrich J, Thurotte A, Schneider D. Specific interaction of IM30/Vipp1 with cyanobacterial and chloroplast membranes results in membrane remodeling and eventually in membrane fusion. Biochim Biophys Acta Biomembr 2017; 1859:537–549
17. Ravi J, Anantharaman V, Aravind L, et al. Variations on a theme: evolution of the phage-shock-protein system in Actinobacteria. Antonie Van Leeuwenhoek 2018; 111:753–760
18. Datta P, Ravi J, Guerrini V, et al. The Psp system of Mycobacterium tuberculosis integrates envelope stress-sensing and envelope-preserving functions. Mol. Microbiol. 2015; 97:408–422
19. Brissette JL, Weiner L, Ripmaster TL, et al. Characterization and sequence of the Escherichia coli

stress-induced psp operon. J. Mol. Biol. 1991; 220:35–48
20. Huvet M, Toni T, Sheng X, et al. The evolution of the phage shock protein response system: interplay between protein function, genomic organization, and system function. Mol. Biol. Evol. 2011; 28:1141–1155
21. Mascher T, Zimmer SL, Smith T-A, et al. Antibiotic-inducible promoter regulated by the cell envelope stress-sensing two-component system LiaRS of Bacillus subtilis. Antimicrob. Agents Chemother. 2004; 48:2888–2896
22. Horinouchi T, Suzuki S, Hirasawa T, et al. Phenotypic convergence in bacterial adaptive evolution to ethanol stress. BMC Evol. Biol. 2015; 15:180
23. Huvet M, Toni T, Tan H, et al. Model-based evolutionary analysis: the natural history of phage-shock stress response. Biochem. Soc. Trans. 2009; 37:762–767
24. Vothknecht UC, Otters S, Hennig R, et al. Vipp1: a very important protein in plastids?! J. Exp. Bot. 2012; 63:1699–1712
25. Bultema JB, Fuhrmann E, Boekema EJ, et al. Vipp1 and PspA: Related but not twins. Commun Integr Biol 2010; 3:162–165
26. Westphal S, Heins L, Soll J, et al. Vipp1 deletion mutant of Synechocystis: a connection between bacterial phage shock and thylakoid biogenesis? Proc. Natl. Acad. Sci. U.S.A. 2001; 98:4243–4248
27. Heidrich J, Wulf V, Hennig R, et al. Organization into Higher Ordered Ring Structures Counteracts Membrane Binding of IM30, a Protein Associated with Inner Membranes in Chloroplasts and Cyanobacteria. J. Biol. Chem. 2016; 291:14954–14962
28. Theis J, Gupta TK, Klingler J, et al. VIPP1 rods engulf membranes containing phosphatidylinositol phosphates. Scientific Reports 2019; 9:8725
29. Brissette JL, Russel M, Weiner L, et al. Phage shock protein, a stress protein of Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 1990; 87:862–866
30. Aseeva E, Ossenbühl F, Eichacker LA, et al. Complex formation of Vipp1 depends on its alpha-helical PspA-like domain. J. Biol. Chem. 2004; 279:35535–35541
31. Jovanovic G, Mehta P, McDonald C, et al. The N-terminal amphipathic helices determine regulatory and effector functions of phage shock protein A (PspA) in Escherichia coli. J. Mol. Biol. 2014; 426:1498–1511
32. Ranea JAG, Sillero A, Thornton JM, et al. Protein superfamily evolution and the last universal common ancestor (LUCA). J. Mol. Evol. 2006; 63:513–525
33. Flores-Kim J, Darwin AJ. Interactions between the Cytoplasmic Domains of PspB and PspC Silence the Yersinia enterocolitica Phage Shock Protein Response. J. Bacteriol. 2016; 198:3367–3378
34. Joly N, Engl C, Jovanovic G, et al. Managing membrane stress: the phage shock protein (Psp) response, from molecular mechanisms to physiology. FEMS Microbiol. Rev. 2010; 34:797–827
35. Domínguez-Escobar J, Wolf D, Fritz G, et al. Subcellular localization, interactions and dynamics of the phage-shock protein-like Lia response in Bacillus subtilis. Mol. Microbiol. 2014; 92:716–732
36. Manganelli R, Gennaro ML. Protecting from Envelope Stress: Variations on the Phage-Shock-Protein Theme. Trends Microbiol. 2017; 25:205–216
37. Otters S, Braun P, Hubner J, et al. The first α-helical domain of the vesicle-inducing protein in plastids 1 promotes oligomerization and lipid binding. Planta 2013; 237:529–540
38. Helmann JD. Bacillus subtilis extracytoplasmic function (ECF) sigma factors and defense of the cell envelope. Curr. Opin. Microbiol. 2016; 30:122–132
39. Eiamphungporn W, Helmann JD. The Bacillus subtilis sigma(M) regulon and its contribution to cell envelope stress responses. Mol. Microbiol. 2008; 67:830–848
40. Jordan S, Junker A, Helmann JD, et al. 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. 2006; 188:5153–5166
41. van der Laan M, Urbanus ML, Ten Hagen-Jongman CM, et al. A conserved function of YidC in the biogenesis of respiratory chain complexes. Proc. Natl. Acad. Sci. U.S.A. 2003; 100:5801–5806
42. Maxson ME, Darwin AJ. Multiple promoters control expression of the Yersinia enterocolitica phage-shock-protein A (pspA) operon. Microbiology (Reading, Engl.) 2006; 152:1001–1010
43. Weiner L, Model P. Role of an Escherichia coli stress-response operon in stationary-phase survival. Proc. Natl. Acad. Sci. U.S.A. 1994; 91:2191–2195
44. Jovanovic G, Lloyd LJ, Stumpf MPH, et al. Induction and function of the phage shock protein extracytoplasmic stress response in Escherichia coli. J. Biol. Chem. 2006; 281:21147–21161
45. Mitchell AM, Silhavy TJ. Envelope stress responses: balancing damage repair and toxicity. Nat. Rev. Microbiol. 2019; 17:417–428
46. Engl C, Beek AT, Bekker M, et al. Dissipation of proton motive force is not sufficient to induce the phage shock protein response in Escherichia coli. Curr. Microbiol. 2011; 62:1374–1385
47. Weiner L, Brissette JL, Model P. 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. 1991; 5:1912–1923
48. Jovanovic G, Engl C, Buck M. Physical, functional and conditional interactions between ArcAB and phage shock proteins upon secretin-induced stress in Escherichia coli. Mol. Microbiol. 2009; 74:16–28
49. Darwin AJ. The phage-shock-protein response. Mol. Microbiol. 2005; 57:621–628
50. Dworkin J, Jovanovic G, Model P. The PspA protein of Escherichia coli is a negative regulator of sigma(54)-dependent transcription. J. Bacteriol. 2000; 182:311–319
51. Elderkin S, Bordes P, Jones S, et al. Molecular determinants for PspA-mediated repression of the AAA transcriptional activator PspF. J. Bacteriol. 2005; 187:3238–3248
52. Yamaguchi S, Reid DA, Rothenberg E, et al. Changes in Psp protein binding partners, localization and behaviour upon activation of the Yersinia enterocolitica phage shock protein response. Mol. Microbiol. 2013; 87:656–671
53. Lloyd LJ, Jones SE, Jovanovic G, et al. Identification of a new member of the phage shock protein response in Escherichia coli, the phage shock protein G (PspG). J. Biol. Chem. 2004; 279:55707–55714
54. Green RC, Darwin AJ. PspG, a new member of the Yersinia enterocolitica phage shock protein regulon. J. Bacteriol. 2004; 186:4910–4920
55. Wolf D, Kalamorz F, Wecke T, et al. In-depth profiling of the LiaR response of Bacillus subtilis. J. Bacteriol. 2010; 192:4680–4693
56. Suntharalingam P, Senadheera MD, Mair RW, et al. The LiaFSR system regulates the cell envelope stress response in Streptococcus mutans. J. Bacteriol. 2009; 191:2973–2984
57. Kesel S, Mader A, Höfler C, et al. Immediate and heterogeneous response of the LiaFSR two-component system of Bacillus subtilis to the peptide antibiotic bacitracin. PLoS ONE 2013; 8:e53457
58. Schrecke K, Jordan S, Mascher T. Stoichiometry and perturbation studies of the LiaFSR system of Bacillus subtilis. Mol. Microbiol. 2013; 87:769–788
59. Kleine B, Chattopadhyay A, Polen T, et al. The three-component system EsrISR regulates a cell envelope stress response in Corynebacterium glutamicum. Mol. Microbiol. 2017; 106:719–741
60. Weiner L, Brissette JL, Ramani N, et al. Analysis of the proteins and cis-acting elements regulating the stress-induced phage shock protein operon. Nucleic Acids Res. 1995; 23:2030–2036

61. Srivastava D, Moumene A, Flores-Kim J, et al. Psp Stress Response Proteins Form a Complex with Mislocalized Secretins in the Yersinia enterocolitica Cytoplasmic Membrane. mBio 2017; 8:
62. DeAngelis CM, Nag D, Withey JH, et al. Characterization of the Vibrio cholerae Phage Shock Protein Response. J. Bacteriol. 2019; 201:
63. Anh Le TT, Thuptimdang P, McEvoy J, et al. Phage shock protein and gene responses of Escherichia coli exposed to carbon nanotubes. Chemosphere 2019; 224:461–469
64. Kingston AW, Liao X, Helmann JD. Contributions of the σ(W) , σ(M) and σ(X) regulons to the lantibiotic resistome of Bacillus subtilis. Mol. Microbiol. 2013; 90:502–518
65. Jordan S, Rietkötter E, Strauch MA, et al. LiaRS-dependent gene expression is embedded in transition state regulation in Bacillus subtilis. Microbiology (Reading, Engl.) 2007; 153:2530–2540
66. Zhang L, Kato Y, Otters S, et al. Essential role of VIPP1 in chloroplast envelope maintenance in Arabidopsis. Plant Cell 2012; 24:3695–3707
67. Armstrong RM, Adams KL, Zilisch JE, et al. Rv2744c Is a PspA Ortholog That Regulates Lipid Droplet Homeostasis and Nonreplicating Persistence in Mycobacterium tuberculosis. J. Bacteriol. 2016; 198:1645–1661
68. Castelle CJ, Banfield JF. Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. Cell 2018; 172:1181–1197
69. Hug LA, Baker BJ, Anantharaman K, et al. A new view of the tree of life. Nat Microbiol 2016; 1:16048
70. Darwin AJ, Miller VL. 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. 2001; 39:429–444
71. Maxson ME, Darwin AJ. PspB and PspC of Yersinia enterocolitica are dual function proteins: regulators and effectors of the phage-shock-protein response. Mol. Microbiol. 2006; 59:1610–1623
72. Becker LA, Bang I-S, Crouch M-L, et al. Compensatory role of PspA, a member of the phage shock protein operon, in rpoE mutant Salmonella enterica serovar Typhimurium. Mol. Microbiol. 2005; 56:1004–1016
73. Aravind L, Ponting CP. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. FEMS Microbiol Lett 1999; 176:111–116
74. Caillat C, Maity S, Miguet N, et al. The role of VPS4 in ESCRT-III polymer remodeling. Biochem. Soc. Trans. 2019; 47:441–448
75. Schöneberg J, Pavlin MR, Yan S, et al. ATP-dependent force generation and membrane scission by ESCRT-III and Vps4. Science 2018; 362:1423–1428
76. Weiss P, Huppert S, Kölling R. Analysis of the dual function of the ESCRT-III protein Snf7 in endocytic trafficking and in gene expression. Biochem. J. 2009; 424:89–97
77. Godinho RM da C, Crestani J, Kmetzsch L, et al. The vacuolar-sorting protein Snf7 is required for export of virulence determinants in members of the Cryptococcus neoformans complex. Sci Rep 2014; 4:6198
78. Webster BM, Colombi P, Jäger J, et al. Surveillance of nuclear pore complex assembly by ESCRT-III/Vps4. Cell 2014; 159:388–401
79. Flores-Kim J, Darwin AJ. Phage shock protein C (PspC) of Yersinia enterocolitica is a polytopic membrane protein with implications for regulation of the Psp stress response. J. Bacteriol. 2012; 194:6548–6559
80. Gueguen E, Savitzky DC, Darwin AJ. Analysis of the Yersinia enterocolitica PspBC proteins defines functional domains, essential amino acids and new roles within the phage-shock-protein response. Mol. Microbiol. 2009; 74:619–633

81. Wallrodt I, Jelsbak L, Thomsen LE, et al. Removal of the phage-shock protein PspB causes reduction of virulence in Salmonella enterica serovar Typhimurium independently of NRAMP1. J. Med. Microbiol. 2014; 63:788–795
82. Burki F, Shalchian-Tabrizi K, Pawlowski J. Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. Biol. Lett. 2008; 4:366–369
83. Hampl V, Hug L, Leigh JW, et al. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic 'supergroups'. Proc. Natl. Acad. Sci. U.S.A. 2009; 106:3859–3864
84. Osadnik H, Schöpfel M, Heidrich E, et al. PspF-binding domain PspA1-144 and the PspA·F complex: New insights into the coiled-coil-dependent regulation of AAA+ proteins. Mol. Microbiol. 2015; 98:743–759
85. Tang S, Henne WM, Borbat PP, et al. Structural basis for activation, assembly and membrane binding of ESCRT-III Snf7 filaments. Elife 2015; 4:
86. Zhang L, Kondo H, Kamikubo H, et al. VIPP1 Has a Disordered C-Terminal Tail Necessary for Protecting Photosynthetic Membranes against Stress. Plant Physiol. 2016; 171:1983–1995
87. Anantharaman V, Aravind L. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome Biol. 2003; 4:R11
88. Brosch R, Gordon SV, Marmiesse M, et al. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A 2002; 99:3684–3689
89. Anantharaman V, Koonin EV, Aravind L. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol 2001; 307:1271–1292
90. Kaur G, Burroughs AM, Iyer LM, et al. Highly regulated, diversifying NTP-dependent biological conflict systems with implications for the emergence of multicellularity. Elife 2020; 9:
91. Wollert T, Wunder C, Lippincott-Schwartz J, et al. Membrane scission by the ESCRT-III complex. Nature 2009; 458:172–177
92. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 2017; 541:353–358
93. Kaur G, Iyer LM, Subramanian S, et al. Evolutionary convergence and divergence in archaeal chromosomal proteins and Chromo-like domains from bacteria and eukaryotes. Sci Rep 2018; 8:6196
94. Burroughs AM, Aravind L. RNA damage in biological conflicts and the diversity of responding RNA repair systems. Nucleic Acids Res 2016; 44:8525–8555
95. Delahay RM, Shaw RK, Elliott SJ, et al. Functional analysis of the enteropathogenic Escherichia coli type III secretion system chaperone CesT identifies domains that mediate substrate interactions. Mol Microbiol 2002; 43:61–73
96. Lee YH, Kingston AW, Helmann JD. Glutamate dehydrogenase affects resistance to cell wall antibiotics in Bacillus subtilis. J. Bacteriol. 2012; 194:993–1001
97. Jovanovic G, Mehta P, Ying L, et al. Anionic lipids and the cytoskeletal proteins MreB and RodZ define the spatio-temporal distribution and function of membrane stress controller PspA in Escherichia coli. Microbiology (Reading, Engl.) 2014; 160:2374–2386
98. Lee M-J, Huang C-Y, Sun Y-J, et al. Cloning and characterization of spermidine synthase and its implication in polyamine biosynthesis in Helicobacter pylori strain 26695. Protein Expr. Purif. 2005; 43:140–148
99. Li B, Kurihara S, Kim SH, et al. A polyamine-independent role for S-adenosylmethionine decarboxylase. Biochem J 2019; 476:2579–2594
100. Cerutti ML, Otero LH, Smal C, et al. Structural and functional characterization of a cold adapted

TPM-domain with ATPase/ADPase activity. J. Struct. Biol. 2017; 197:201–209
101. Wegener KM, Bennewitz S, Oelmüller R, et al. The Psb32 protein aids in repairing photodamaged photosystem II in the cyanobacterium Synechocystis 6803. Mol Plant 2011; 4:1052–1061
102. Sirpiö S, Allahverdiyeva Y, Suorsa M, et al. TLP18.3, a novel thylakoid lumen protein regulating photosystem II repair cycle. Biochem J 2007; 406:415–425
103. Wiseman B, Nitharwal RG, Fedotovskaya O, et al. Structure of a functional obligate complex III2IV2 respiratory supercomplex from Mycobacterium smegmatis. Nat Struct Mol Biol 2018; 25:1128–1136
104. Ballas SK, Mohandas N, Marton LJ, et al. Stabilization of erythrocyte membranes by polyamines. Proc. Natl. Acad. Sci. U.S.A. 1983; 80:1942–1946
105. Fiuza M, Letek M, Leiba J, et al. Phosphorylation of a novel cytoskeletal protein (RsmP) regulates rod-shaped morphology in Corynebacterium glutamicum. J. Biol. Chem. 2010; 285:29387–29397
106. Eberhardt RY, Bartholdson SJ, Punta M, et al. The SHOCT domain: a widespread domain under-represented in model organisms. PLoS ONE 2013; 8:e57848
107. Miller C, Kong J, Tran TT, et al. Adaptation of Enterococcus faecalis to daptomycin reveals an ordered progression to resistance. Antimicrob. Agents Chemother. 2013; 57:5373–5383
108. Wang L, Xiao S, Chen X, et al. ytiB and ythA Genes Reduce the Uranium Removal Capacity of Bacillus atrophaeus. Int J Mol Sci 2019; 20:
109. Boyle-Vavra S, Yin S, Jo DS, et al. VraT/YvqF is required for methicillin resistance and activation of the VraSR regulon in Staphylococcus aureus. Antimicrob. Agents Chemother. 2013; 57:83–95
110. Gardete S, Wu SW, Gill S, et al. Role of VraSR in antibiotic resistance and antibiotic-induced stress response in Staphylococcus aureus. Antimicrob. Agents Chemother. 2006; 50:3424–3434
111. Wolf D, Domínguez-Cuevas P, Daniel RA, et al. Cell envelope stress response in cell wall-deficient L-forms of Bacillus subtilis. Antimicrob. Agents Chemother. 2012; 56:5907–5915
112. Hauf S, Möller L, Fuchs S, et al. PadR-type repressors controlling production of a non-canonical FtsW/RodA homologue and other trans-membrane proteins. Sci Rep 2019; 9:10023
113. Zhang L, Sakamoto W. Possible function of VIPP1 in maintaining chloroplast membranes. Biochim. Biophys. Acta 2015; 1847:831–837
114. Leipe DD, Wolf YI, Koonin EV, et al. Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol 2002; 317:41–72
115. Lane N, Martin WF. The origin of membrane bioenergetics. Cell 2012; 151:1406–1416
116. Sojo V, Pomiankowski A, Lane N. A bioenergetic basis for membrane divergence in archaea and bacteria. PLoS Biol 2014; 12:e1001926
117. Benson DA, Cavanaugh M, Clark K, et al. GenBank. Nucleic Acids Res. 2013; 41:D36-42
118. Tatusova T, Ciufo S, Federhen S, et al. Update on RefSeq microbial genomes resources. Nucleic Acids Res. 2015; 43:D599-605
119. Wattam AR, Abraham D, Dalay O, et al. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. 2014; 42:D581-591
120. Wattam AR, Davis JJ, Assaf R, et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res. 2017; 45:D535–D542
121. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016; 44:D733-745
122. Federhen S. The NCBI Taxonomy database. Nucleic Acids Res. 2012; 40:D136-143

123. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402
124. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009; 10:421
125. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005; 33:W244-248
126. Nielsen H. Predicting Secretory Proteins with SignalP. Methods Mol. Biol. 2017; 1611:59–73
127. Krogh A, Larsson B, von Heijne G, et al. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 2001; 305:567–580
128. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 2004; 338:1027–1036
129. Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res. 2007; 35:W429-432
130. Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. Nucleic Acids Res. 2008; 36:W197-201
131. Mistry J, Chuguransky S, Williams L, et al. Pfam: The protein families database in 2021. Nucleic Acids Res 2021; 49:D412–D419
132. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011; 39:W29-37
133. Lassmann T. Kalign 3: multiple sequence alignment of large data sets. Bioinformatics 2019;
134. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792–1797
135. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS ONE 2010; 5:e9490
136. Chang W, Cheng J, Allaire JJ, et al. shiny: Web Application Framework for R. 2019;
137. Winston Chang RCT Joe Cheng, JJ Allaire, Yihui Xie, Jonathan McPherson, RStudio, jQuery Foundation, jQuery contributors, jQuery UI contributors, Mark Otto, Jacob Thornton, Bootstrap contributors, Twitter, Inc, Alexander Farkas, Scott Jehl, Stefan Petre, Andrew Rowls, Dave Gandy, Brian Reavis, Kristopher Michael Kowal, es5-shim contributors, Denis Ineshin, Sami Samhuri, SpryMedia Limited, John Fraser, John Gruber, Ivan Sagalaev. Web Application Framework for R. 2020;
138. R Core Team. R: A Language and Environment for Statistical Computing. 2020;
139. RStudio Team. RStudio: Integrated Development Environment for R. 2020;
140. Mike Bostock CY Kerry Rodden, Kevin Warne, Kent Russell, Florian Breitwieser. Sunburst 'Htmlwidget'. 2020;
141. Hao ZhuThomas Travison DM Timothy Tsai, Will Beasley, Yihui Xie, GuangChuang Yu, Stéphane Laurent, Rob Shepherd, Yoni Sidi, Brian Salzer, George Gui, Yeliang Fan. Construct Complex Table with 'kable' and Pipe Syntax. 2020;
142. Lex A, Gehlenborg N, Strobelt H, et al. UpSet: Visualization of Intersecting Sets. IEEE Transactions on Visualization and Computer Graphics 2014; 20:1983–1992
143. Enrico Bonatesta UB Christoph Horejs-Kainrath. Multiple Sequence Alignment. 2020;
144. Emmanuel Paradis D de V Simon Blomberg, Ben Bolker, Joseph Brown, Santiago Claramunt, Julien Claude, Hoa Sien Cuong, Richard Desper, Gilles Didier, Benoit Durand, Julien Dutheil, RJ Ewing, Olivier Gascuel, Thomas Guillerme, Christoph Heibl, Anthony Ives, Bradley Jones, Franz Krah, Daniel Lawson, Vincent Lefort, Pierre Legendre, Jim Lemon, Guillaume Louvel, Eric Marcon, Rosemary McCloskey, Johan Nylander, Rainer Opgen-Rhein, Andrei-Alin Popescu, Manuela

Royer-Carenzi, Klaus Schliep, Korbinian Strimmer. Analyses of Phylogenetics and Evolution. 2020;
145. Fellows I. Word Clouds. 2018;
146. Dawei Lang GC. Create Word Cloud by htmlwidget. 2018;
147. Jake Conway NG. A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets. 2019;
148. Müller K. here: A Simpler Way to Find Your Files. 2017;
149. Wickham H, Mara Averick HY Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo. Welcome to the tidyverse. Journal of Open Source Software 2019; 4:1686
150. Martin Morgan MR. Access the Bioconductor Project Package Repository. 2019;
151. Perry M. Flexible Heatmaps for Functional Genomics and Sequence Features. 2016;
152. Xie Y, Allaire JJ, Grolemund G. R Markdown: The Definitive Guide. 2018;
153. Allaire JJ, Xie, Y, McPherson J, et al. rmarkdown: Dynamic Documents for R. 2020;