# Phage-shock-protein (Psp) Envelope Stress Response: Evolutionary History & Discovery of Novel Players

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; [2]Public Health Research Institute, Rutgers University, Newark, NJ; [3]National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD.

*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) stress-response system protects bacteria from envelope stress and stabilizes the cell membrane. Recent work from our group suggests that the psp systems have evolved independently in distinct Gram-positive and Gram-negative bacterial clades to effect similar stress response functions. Despite the prevalence of the key effector, PspA, and the functional Psp system, the various genomic contexts of Psp proteins, as well as their evolution across the kingdoms of life, have not yet been characterized. We have developed a computational pipeline for comparative genomics and protein sequence-structure-function analyses to identify sequence homologs, phyletic patterns, domain architectures, gene neighborhoods, and evolution of the candidates across the tree of life. This integrative pipeline enabled us to make several new discoveries, including the truly universal nature of PspA and the ancestry of the PspA/Snf7 dating all the way back to the Last Universal Common Ancestor. Using contextual information from conserved gene neighborhoods and their domain architectures, we delineated the phyletic patterns of all the Psp members. Next, we systematically identified all possible 'flavors' and genomic neighborhoods of the Psp systems. Finally, we traced their evolution, leading us to several interesting findings of their occurrence and co-migration that together point to the functions and roles of Psp in stress-response systems that are often lineage-specific. Conservation of the Psp systems across bacterial phyla emphasizes the established importance of this stress response system in prokaryotes, while the modularity in various lineages is indicative of adaptation to bacteria-specific cell-envelope structures, lifestyles, and adaptation strategies. We also developed an interactive web application that hosts all the data and results in this study that researchers can explore (https://jravilab.shinyapps.io/psp-evolution).

## Keywords

Phage shock protein response | Psp | Envelope stress response | Phylogeny and evolution | Domain architectures | Genomic contexts/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 the cytoskeletal proteins [2]. Even though the precise composition and structure of the membranes and associated lipids and proteins could vary widely, most of the membrane functions enumerated above are conserved across the tree of life [1]. Cellular membranes are constantly adapting and surviving in their ever-changing environments, especially when subject to various kinds of surface stress. When the membrane integrity is compromised due to damage, a specialized suite of lipids and proteins participate in sensing and responding to the stress [3,4], instigating a chain of events to restore membrane function. Extra-cytoplasmic envelope stress responses are common virulence mechanisms employed by bacterial pathogens [5–7]. Similarly, antibiotics, antimicrobial peptides (natural or synthetic) could trigger the cell envelope stress responses [8,9]. Some of the pathways involved in the restoration of biophysical properties could be conserved. Others, such as the envelope stress response machinery maintaining membrane homeostasis, could comprise both conserved and lineage-specific molecular systems. Achieving similar solutions to broad-spectrum problems such as environmental stress via phenotypic convergence is a common occurrence, even if it's not always paired with genotypic convergence [10]. Therefore, to fully understand the molecular basis of membrane function and restoration, it is essential to map out how these systems have evolved alongside each other on the evolutionary timescale, an aspect that remains unclear.

Multiple protein systems are intrinsically tied to membrane dynamics, remodeling, and stabilization, such as ESCRT systems. One such respondent is Vipp1, a predominantly plant protein with a cyanobacterial origin [11–14]. This peripheral membrane protein is involved in the vesicular transport of lipids and thylakoid biogenesis [12,13,15,16]. This inner membrane protein bears similarity to other bacterial and eukaryotic protein families involved in secretory pathways, protein sorting and transport, and stabilization of bent membranes. Either directly via membrane biogenesis and stabilization, or indirectly via regulatory systems, Vipp1 is known to preserve membrane function under stress. Phylogenetic analyses have revealed that Vipp1 has homology to PspA [11,14,17], a bacterial peripheral-membrane protein with known association to stress response functions. Since the membranes of unicellular organisms interact directly with the external environment and stimuli, sensing of these stress cues and repair/maintenance of membrane integrity becomes paramount. In line with this argument, it has been reported that the primary role of PspA, similar to Vipp1, is the preservation of envelope function in response to a wide range of surface stresses via membrane fusion [15] and prevention of leakage of protons through the membrane [2,18]. These perturbations include changes in membrane potential, Δ*pH* changes [2,18], and PMF [19–22], (l)antibiotic stress [9], heat and osmotic shock [23], and mechanical stress via stored curvature elastic stress [24–27], among others [23].

Structurally, both PspA and Vipp1 contain the *PspA_IM30* domain that has a predominantly coiled-coil structure [11,12,15,28–33] [**Table 1**]. The N-terminal amphipathic helices associate with membranes to sense stored curvature elastic stress [25,32]. The *PspA_IM30* domain is found in bacteria, photosynthetic eukaryotes, and archaea [11,14,34–39] [**Fig. 1**]. Both Vipp1 and PspA are

present in the photosynthetic cyanobacteria [11,15,40], hinting at the origins and phylogenetic migration of the domain over to eukaryotes. Of the two, only PspA retains the stress sensing role [9,18,37,41–43] while Vipp1 is involved in thylakoid biogenesis [12,13,15].

Bacterial PspA has been known to function with cognate partners, typically transcriptional regulators, and membrane proteins. The three best described Psp systems span the dominant bacterial lineages, namely, proteobacteria (*pspF||ABCDE-G* [22,28,35,44–49]), firmicutes (*liaIHGFSR* [36,39,43,50–53]), and actinobacteria (*clgRpspAMN* [37,38]) [**Fig. 1**]. Each of these cognate members serves 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 [20,47,54–59] or the LiaRS two-component system [39,43,50,51,60] have been implicated in direct stress-sensing function independent of PspA. The classical Psp system that was initially discovered in Proteobacteria [28,29], however, is not universal, *i.e.*, the system is not conserved across all bacterial/archaeal clades. Despite the presence of variable genomic contexts, PspA seems to have evolved to function towards the preservation of the cell membrane. Current studies have mostly focused on lineage-specific Psp systems (*e.g.,* Proteobacteria, Firmicutes, Actinobacteria, Plants [5,11,12,14,22,34,37,38,61–63]). 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. The Psp system is a specific envelope stress response system that seems to have converged functionally while retaining a pretty diverse and divergent lineage-specific genotypic signature.

In this paper, we have taken a generalized and global approach while remaining unbiased to 'the known' to delineate all possible occurrences of PspA and its cognate partners, their associated genomic contexts, and phyletic spreads across the bacterial and archaeal kingdoms. These molecular characterization steps would help ascertain the evolutionary and functional significance by answering the following questions: i) 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? ii) Have the Psp proteins evolved to function in other (i.e., not one of the known three) genomic contexts? iii) Are there common patterns/themes that emerge from the functional characterization of these Psp systems, and what do we understand about their lineage-specificity? vi) Finally, is the migration and, thereby, the functional implications of these different Psp members more independent than originally anticipated? To address the conservation of PspA and its potential partners, we conducted an evolutionary study across all the major kingdoms of life [64,65] by querying for Vipp1, PspA, and its many distinct paralogs (across bacterial, plant, and other archaeal and eukaryotic clades [11,12,14,31]), and other PspA-like proteins that function to preserve membranes across the tree of life. We subsequently analyze the different variants of PspA, and PspA-associated proteins (PspBC, PspMN, LiaIGF), based on sequence-structure relationships overlaid on phyletic contexts. The resulting findings help discern the truly universal nature of this envelope stress-response system in an evolutionary light.

# Results and Discussion

To understand the evolution of the Psp system and its ability to function across phyla, we needed to reconcile two seemingly disparate facts: the universality of PspA and the distinct genomic neighborhoods containing PspA [34,38,61] 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 began our analyses by exploring their phyletic spreads [**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 of the constituent domains 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 for the detection of domains, signal peptides, transmembrane regions to help construct the domain architecture of each of the query Psp proteins [See Methods]. Due to the prevalence of transcription factor domains and two-component systems in bacterial phyla (PspF and ClgR with the HTH domains, NtrC-AAA—FIS-HTH, XRE_HTH; LiaRS with HAMP-HISKIN, REC-LYTTR domains **Fig. 1A**), we did not perform dedicated searches for these proteins/domains; instead, we noted their occurrence in the genomic neighborhood of core Psp genes. Where required, as in the case of the Histidine Kinase, we identified the closest orthologs to study the full extent of the diversity.

The section below describes the delineation of the core Psp proteins in proteobacteria, actinobacteria, and firmicutes, in terms of their underlying domain architectures, and their phyletic spreads [**Fig. 1**]. Then, we outlined 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 by determining their domain architecture, genomic contexts, and phyletic spreads [**Fig. 3**]. Finally, we report the findings from our in-depth analysis of domain architectures [**Fig. 4**], genomic contexts [**Fig. 5**], and phyletic spreads to every one of the Psp cognate partners to identify novel components (domains and genomic neighborhoods) of the phage shock protein stress response system across the tree of life [**Fig. 4, 5**]. All these new findings are reconciled in the form of a proximity network of domains with respect to PspA along with their phyletic spreads and frequencies [**Fig. 6**]. All our findings (data, results, and visual summaries) are available for in-depth exploration 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 started with the Psp operon that was identified in *E. coli*, and subsequently, in other proteobacteria [14,28,29,66–68]. We resolved the core members of the operon, PspF||ABC, into their

domain architectures. PspA, PspB, and PspC each comprising their namesake domains, *PspA_IM30* (*IM30* standing for inner membrane 30KDa domain, henceforth called PspA), PspB, and PspC, respectively, and each domain spanning almost the entire protein length [**Fig. 1A**; **Table 1**]. In line with previous studies [46], we identified that PspF, the transcriptional regulator of the Psp operon in *E. coli*, contains an enhancer-binding *NtrC-AAA (AAA-ATPase)* domain in addition to the *FIS-HTH* (helix-turn-helix) DNA-binding domain [**Fig. 1B**].

### The Psp system in actinobacteria

The protein central to the identification of the Psp operon in *M. tuberculosis* (actinobacteria) is PspA (encoded by *rv2744c*). We confirmed that PspA contains the PspA domain, albeit with <30% similarity to the *E. coli* homolog. However, in contrast to a PspF-like transcriptional regulator in *E. coli*, we observed that the *M. tuberculosis* operon comprises an XRE-type HTH transcription factor family member, ClgR, that is unique to actinobacteria [**Fig. 1B**]. Likewise, *M. tuberculosis* membrane protein, Rv2743c, is distinct from its *E. coli* counterparts, PspC and PspB [**Fig. 1B**]. The final member of the four-protein actinobacterial operon, Rv2742c, contains a domain of unknown function, DUF3046 [**Fig. 1B**; **Table 1**]. Despite its requirement for the full activity of the Psp system [37], the molecular nature of the Rv2742c protein remains to be characterized. These findings are consistent with previous studies [37,38,61]. To remain consistent with the nomenclature of Psp cognate proteins, we rename Rv2743c (membrane protein) and Rv2742c (carrying a C-terminal DUF3046) to 'PspM' and 'PspN', respectively [38].

### The Psp system in firmicutes

Finally, we examined the domain architectures of the well-studied Lia operon in *B. subtilis*, revealing some interesting features. We initially identified, in line with recent studies [39,51,60], that: LiaI is a small transmembrane protein; LiaF is a protein with N-terminal transmembrane helices (DUF2157) and a C-terminal globular domain (DUF2154); LiaG contains another domain of unknown function (DUF4097 that is similar to DUF2154) [**Fig. 1B**]. Our analysis additionally showed that the Transmembrane domains containing DUF2157 encompass the LiaI-LiaF-TM domain and those containing the DUF4097/ DUF2154 encompass the Toast_rack domain (see below for details on our newly defined nomenclatures, LiaI-LiaF-TM and Toast_rack) [**Fig. 1B**]. As documented previously [39,43], we found that, in this firmicute operon, the transcriptional response is brought about by the two-component system, LiaRS, comprising the REC-LYTTR and HAMP-HISKIN domains [**Fig. 1B**].

**In summary**, the above data show that, in addition to domain architectures previously described, such as PspA, PspB, PspC in proteobacteria, and the transcription regulators, we have identified novel domain definitions, such as LiaI-LiaF-TM and Toast_rack, in firmicutes [**Fig. 1B**]. Having deconvolved the Psp member proteins of the 'known' systems, we next focused on where these proteins are found 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, the namesake protein of the system. In line 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 1**), occurs in most of the major bacterial and archaeal lineages as 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 were collected and collated. This analysis revealed that the PspA_IM30 (now referred to as PspA) domain is homologous to the eukaryotic Snf7 superfamily [**Fig. 1B**], which is part of the membrane remodeling ESCRT system [69–74]. Further, our findings showed that iterative searches with PspA/Vipp1 and Snf7 recover each other, suggesting that they are indeed part of the same superfamily. Iterative searches with the Snf7 domain recovered several non-plant eukaryotes and archaeal homologs [**Fig. 1B**]. Taken together, the PspA and Snf7 superfamilies are widely prevalent, extending to almost every distinct phyletic lineage spanning all three kingdoms [**Fig. 1C**, with the PspA branch conserved in Bacteria with transfers to archaea and eukaryotes, and the Snf7 branch conserved in Archaea and Eukaryotes with transfers to bacteria. Taken together, we found that PspA/Snf7 is the only 'universal protein' in the Psp system present in all three kingdoms: bacteria, archaea, and eukaryota [**Fig. 1B, Table S1**].

## LiaI-LiaF-TM and Toast_rack

Next, we tracked the presence of LiaF, LiaG, and LiaI across the tree of life using their whole protein sequences and their constituent domains, including our new domain definitions, LiaI-LiaF-TM and Toast_rack. [**Fig. 1B**]. Specifically, we used PSI-BLAST searches from the three sub-sequences of the whole proteins, N-terminal transmembrane 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 transmembrane and C-terminal globular regions of the LiaF protein, respectively. In fact, 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 Toast_rack (PDB: *4QRK*; **Table 1**; webapp). We will, therefore, collectively refer to these three domains of unknown function as the '**Toast_rack**' domain [**Fig. 1A**; **Table 1**]. Likewise, the high level of similarity between the 4TM (four transmembrane) regions of LiaI and the N-terminal domain, DUF2157, of LiaF led us to rename the *4TM* region (previously referred to as Toast_rack_N, Pfam: PF17115) to be '**LiaI-LiaF-TM**' [**Fig. 1B**; **Table 1**]. Toast_rack-containing proteins are panbacterial with transfers to archaea and eukaryotes [**Fig. 1B**]. We found a few metazoan, fungal, and SAR-group eukaryotic genomes that carry this globular *Toast_rack* domain as well [**Fig. 1B**]. 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**]. Taken together, the results of our analyses define two new domains: LiaI-LiaF-TM and Toast_rack.

## PspC

PspC, a membrane protein first identified in the proteobacterial Psp system, is critical in sensing the membrane stress response and restoring envelope integrity [54,67,75]. A few recent studies have shown that PspC may function independently of PspA in some bacterial species in response to membrane stressors such as secretin toxicity [54,56,67,75,76]. Our analysis showed that PspC has

two transmembrane helices, the first of them being a cryptic TM. A conserved amino acid residue, R, is found in between the two TM regions, suggesting a role in sensing ['MSA' tab in the webapp]. The PspC domain is panbacterial and is found in a few archaeal clades [**Fig. 1C**].

## PspB

PspB is another membrane protein, often found to be part of the proteobacterial Psp operon, and implicated in stress sensing and virulence [67,76,77]. PspB has an N-terminal anchoring transmembrane helix followed by an alphahelical globular domain. We found that PspB is almost completely restricted to Proteobacteria with a few transfers and with fusions only to PspC [**Fig. 1B**; See the section below on *PspC domain architectures* for further details].

## PspM/PspN

Based on our previous [37,38,61] and this current study, we found no discernible homologs for these proteins (and constituent domains) outside the phylum actinobacteria [**Fig. 1C**].

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 evident from 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 and Nocardia, Actinopolysporales, Frankiales, Micromonosporales, Nakamurellales and Pseudonocardiales [**Fig. 1B**] in accordance with our previous findings [37,38,61].

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 1**]. As observed in our previous studies [37,38], we found that the DUF3046 domain is widely prevalent across actinobacteria, but not present in other phyla [**Fig. 1C**]. Moreover, the *M. tuberculosis* genome shows the existence of a second copy of this domain DUF3046, only four genes downstream (*rv2738c*). We found that DUF3046 is alpha-helical with mostly conserved T and C residues as seen in the multiple sequence alignment.

We examined the DUF3046 homologs in more detail and searched for the connection between the two mycobacterial copies of the domain using the genome sequences (rather than translated protein ORFs) followed by sequence alignment reveals. This analysis revealed a few interesting features [**Fig. 1B**]. First, we found that the DUF3046 domain that is widespread across actinobacteria is closer to the short downstream protein, Rv2738c, rather than the C-terminus of the fourth member of the *M. tuberculosis* operon, PspN. We thus infer that Rv2738c, rather than PspN, contains the ancestral copy of the DUF3046. The DUF3046 domain that is part of the open reading frame (ORF) in PspN is likely a duplicated copy of the domain that jumped into the PspN ORF within mycobacteria, especially the *M. tuberculosis* complex. Moreover, unlike the pan-actinobacterial DUF3046, the N-terminal portion of PspN, PspN_N, is conserved only within *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 careful multiple sequence alignments [**Fig. 1B**; webapp] and delineated the phyletic spreads of PspA and its cognate partner proteins/domains, PspBC, PspMN (DUF3046), LiaIFG (LiaI-LiaF-TM, Toast_rack) across the tree of life [**Fig. 1C**]. While the membrane domains, PspB, PspM, LiaI-LiaF-TM, are highly lineage-specific (restricted to proteobacteria, actinobacteria, and firmicutes, respectively), the other two domains, PspC and Toast_rack, are much more widely-spread [**Fig. 1C**]. In addition to being present in a wide range of bacterial phyla, PspC is present in other archaeal clades and Toast_rack is present in both 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 that we found in the proximity of these Psp systems, we next carried out a detailed analysis of domain architectures, genomic contexts, and the phyletic spreads of each of these homologous 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.

### 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 of the bacterial and archaeal clades contain PspA homologs [**Fig. 1B**]. Of the eukaryotes, only the bikonta (eukaryotes containing two flagella; comprising the SAR/HA supergroup and excavata) have homologs with the PspA domain [**Fig. 2; 1B**; [78,79]].

*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 [69–74]. Based on the curated search results of PspA homologs [See Methods; 'Data' tab of webapp], we uncovered a distinct signature revealing proteins belonging to the eukaryotic Snf7 family among the remote homologs of PspA [**Fig. 2**]. 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 intertwined 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; [64,65]] as well as different domain architectures [**Fig. 2, 3**; detailed in the next section]. We also included PspA/Snf7 paralogs, when available. We further 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. 2**; **Table 1** [80,81]]. With the representative PspA/Snf7 proteins spanning the three kingdoms, we performed a carefully-curated multiple sequence alignment with the selected ~450 PspA/Snf7 homologs [webapp; See Methods]. Finally, we used this multiple sequence alignment to generate a PspA/Snf7 phylogenetic tree [**Fig. 2**] to trace the evolutionary history of the superfamily. The multiple sequence alignment of PspA/Snf7 also highlights variation in

the position, length, and nature of the alpha-helices, *e.g.,* four helices in most bacterial and archaeal lineages [webapp]. Of particular 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 the source of the plant variant, Vipp1 [31,82].

To complete our understanding of the phylogeny, we used the sequence alignment of a representative subset of PspA/Snf7 containing species [**Fig. 2**] to construct a phylogenetic tree [**Fig. 2**; See Methods]. The Snf7 homologs serve as the out-group in our PspA tree. The first striking finding is that several bacterial (such as proteobacteria, actinobacteria, firmicutes) and archaeal clades, along with presumably their ancestral cyanobacterial relatives [**Fig. 2**] self-organize into distinct clusters (each major clade represented by a different color), 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 key observations from the tree (and the underlying sequence alignment) [**Fig. 2**] 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 suggesting the PspA-Vipp1 connection in cyanobacteria; and iii) The *Snf7* domain-containing homologs from archaea and eukaryota form a well-defined branch away from the *PspA* containing branches.

**In summary**, homologs from the PspA and Snf7 superfamilies result in a noteworthy finding that the helical membrane-stress response domain, *PspA*, is likely ancient with the earliest copy existing in the last universal common ancestor (LUCA). We have constructed a representative phylogenetic tree of PspA/Snf7 (**Fig. 2**) based on a carefully hand-curated multiple sequence alignment (webapp). Next, we proceeded to characterize 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 diverse genomes.

### Domain Architectures

Our comprehensive searches starting with eight representative PspA proteins revealed that most PspA homologs (>98%; ~2500 homologs) do not show much variation in terms of their underlying domain architecture. In most lineages, PspA homologs only contain the characteristic PspA (PspA_IM30) domain without additional fusions [**Fig. 3; Table S1**]. The remaining small fraction of homologs (<2%), they contain either repeated PspA domains or fusions with domains other than PspA [**Fig. 3**]. For instance, cyanobacterial PspA homologs show some interesting variations: a few have dyads or triads of PspA domains stringed together [**Fig. 3**; *e.g.,* BAG06017.1] while others have an additional NLPC peptidase domain at the N-terminal end, suggesting a cell wall peptidase role for these inner membrane PspA homologs [**Fig. 3; Table S1**; AFZ52345.1*;* [83]].

Similar to the 98% of PspA homologs, a search for the related superfamily, Snf7, again revealed minimal variation in domain architecture, with fusions only in eukaryotes. The homologs carry one or more copies of the Snf7 domain (only found in eukaryota and some archaea) [Fig. 3b; Table A2], suggestive of a conventional role in oligomerization and membrane stabilization attributed to the eukaryotic ESCRT systems [81].

### *Paralogs*

We found that in genomes that contain multiple copies of PspA, the paralogs do not always have the same domain architecture and/or genomic contexts (discussed further below), *e.g.,* firmicutes, spirochaetes [**Fig. 3**; **Table S1**; 'Phylogeny –> Paralog' tab of the webapp]. We noted a special case in cyanobacteria with 2-3 paralogs, two of which are adjacent proteins in the genome [**Fig. 3**; **Table S1**; BAG06017.1]. This 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 the adjoining PspA proteins in cyanobacteria, a few archaeal and eukaryotic species have adjoining Snf7 proteins (CBY21170.1*; Oikopleura dioica*) suggestive of local gene duplication. The webapp ('Phylogeny' –> 'Paralog') contains the full list of PspA/Snf7 paralogs along with their domain architectures and genomic contexts with likely insights on their shared history, gene duplication vs horizontal gene transfer.

The domain architectures of PspA and Snf7, although limited in their repertoire, are prevalent in all three kingdoms of life, sometimes containing multiple copies within a genome. The ubiquity of the PspA/Snf7 superfamily indicates their inherent utility as maintainers of envelope integrity.

## Novel variations of known genomic contexts

To understand their multifarious roles better, we next explored the gene neighborhoods in the vicinity of each of the Psp members [**Fig. 3**; **Table S1**]. The genomic context analysis uncovered many distinct and novel themes of the psp system; the major contextual themes are summarized below. The full list of genomic contexts and their phyletic spreads are available in the webapp ('Data' and 'Genomic Contexts' tabs, PspA/Snf7 selection in the dropdown).

### *PspFABC operon*

We found that PspA, PspB, and PspC appear in an operon in alpha, delta, and gammaproteobacteria, and a few spirochaetes and nitrospirae [**Fig. 3**; **Table S1**]. We also found a few variations to this 'classic' 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 S1**; *e.g.,* ANW39986.1, *E. coli*], while others have multiple PspB copies in the operon [**Fig. 3**; **Table S1**; *e.g.,* AOL22920.1, *Erythrobacter litoralis*]. PspD occurs along with this operon only in gammaproteobacteria [**Fig. 3**; **Table S1**; *e.g.,* ANW39986.1, *E. coli*]. The transcription regulator PspF (NtrC-AAA and HTH) occurs with a common promoter with the PspABC operon and in a few gammaproteobacteria with only PspA. The whole operon has been transferred to a few spirochaetes (*e.g.,* Spico_0138) with additional variations involving PspB, PspC fusions, and Toast_rack containing proteins.

We found that a few operons in gammaproteobacteria (*e.g.,* Entas_2342 AEN65073.1) have an ACT and PAS domain followed by the NtrC-AAA, and in these cases, two proteins, a GTPase and a membrane-associated IIGP, have been inserted into the operon between the NtrC-AAA protein and PspABCD. A few of these have an additional canonical PspF in the operon. In some alphaproteobacteria, the PspC in PspFABC has been replaced by multiple Toast_rack containing proteins [**Fig. 3**; **Table S1**; *e.g.,* ANQ40502.1, *Gluconobacter oxydans*].

*Classical-AAA-ATPase*

Snf7 and the VPS4-like AAA-ATPase (with an N-terminal MIT domain, TPR-like) are known to occur together in archaea in the ESCRT system (*e.g.,* OLS27540.1) [84]. We found additional divergent Snf7 genes in the operon, including fusion to HTH (*e.g.,* OLB70630.1) in the archaeal eukaryotic clade, with a few bacterial transfers, mostly without the ATPase in the operon [**Table S1**]. The corresponding VPS4-like AAA-ATPase transferred from archaea to bacteria in cyanobacteria, bacteroidetes, verrucomicrobia, nitrospirae and planctomycetes (*e.g.,* ACB74714.1 *Opitutus terrae*) have TPR repeats followed by TM region, instead of Snf7. Further, our analysis showed that the bacterial eukaryotic PspA also occurs with an AAA-ATPase, in various bacterial clades. This gene dyad also occurs with either a previously unidentified signal peptide and transmembrane helix containing protein with a divergent Snf7 domain (OGG56892.1) or other coiled-coil or alphahelical domain-containing proteins. Both PspA and the membrane-associated Snf7 along with the AAA-ATPase occur in a longer operon with ABC-ATPase, ABC Transporter, and a PBPB-OmpA containing protein.

*Membrane dynamics with PspA and ATPase*

Snf7 and VPS4-ATPase have previously been shown to be involved in the ESCRT-III mediated membrane remodeling in archaea [69–71,73]. Here, we found a similar operon of PspA and ATPase in bacteria, suggesting that a similar interaction takes place in bacteria, where PspA and ATPase are involved in the bacterial membrane dynamics [**Fig. 3**]. Sporadically in bacteria, the PspA–AAA-ATPase module is also involved in a novel complex with both cell-surface/membrane-linked transporter and cytoplasmic PspA/Snf7 elements. The OmpA domain interacting with extracellular peptidoglycan and the ABC transporter transport a ligand in response to stress, which in turn interacts with the PspA AAA-ATPase module. Such a system could link extracellular events with PspA internally.

*Novel PspAA association*

We identified a new component in the proximity of PspA, based on the PspA neighborhood analysis. A protein with a novel domain with an HHEH structure, present in euryarchaeota, thaumarchaeota, actinobacteria, chloroflexi, firmicutes, and a few alpha- and gammaproteobacteria), occurs as a two-gene cluster with PspA [**Fig. 3**; **Table S1**; *e.g.,* MA_1460; AAM04874.1, Methanosarcina]. This domain, mostly occurring by itself, is fused to an N-terminal PspA in actinobacteria and chloroflexi [**Fig. 3**; **Table S1**; *e.g.,* ACU53894.1, Acidimicrobium]. We call this protein **PspAA** (for **PspA-Associated** protein). This PspA–PspAA dyad occurs by itself or in two other contexts: 1) with a two-component system (discussed below; CAB51252.1, Streptomyces) or 2) another dyad comprising a membrane-associated Metallopeptidase and protein with a novel domain, **Tfu_1009** with a few variants in not just actinobacteria and archaea, but firmicutes, chloroflexi, and a few other bacterial lineages as well [**Fig. 3**; **Table S1**; ABW11964.1, Frankia; AKB54760.1, Methanosarcina; AKX93460.1, Moorella].

*PspA with PspM (ClgRPspAMN) or Thioredoxin*

PspA occurs as a dyad with Thioredoxin in cyanobacteria and actinobacteria [38] [**Fig. 3**]. Analysis of this dyad showed that the actinobacterial PspA is typified by RsmP from *Corynebacterium testudinoris* (AKK09942.1; **Fig. 3**; **Table S1**), phosphorylation of which has been shown to regulate the rod-shaped morphology in *Corynebacterium glutamicum* [85]. This actinobacterial PspA homolog

belongs to an RsmP family cluster comprising mostly of rod-shaped actinobacteria that have PspA occurring either with **Thioredoxin** or with a **ClgR-HTH**, **PspM**, and an occasional low complexity protein (AOS62694.1, Actinoalloteichus) [**Fig. 3**; **Table S1**]. The Corynebacterial members of the family have paralogs with both versions of PspA contexts ['Paralog' tab of the webapp]. The association of the ClgR-HTH-PspM and Thioredoxin and their mutual exclusion in the operon suggest that they act as repressors in the case of the HTH and redox agent in the case of the thioredoxin to control the PspA [**Fig. S1**]. The association of ClgR-HTH-PspAM is also confined to this RsmP family, suggesting that these are also determinants of rod-shaped morphology of the cell. The PspN presence in the immediate operon of ClgR-HTH-PspAM is limited to a few Mycobacteria (CCP45543.1, *M. tuberculosis H37Rv*), which have an N-terminal PspN-N (christened below) and C-terminal DUF3046. In the remaining Clgr_HTH-PspAM operons, the DUF3046 domain has drifted apart and lacks the N-terminal domain fusion (ABK71106.1, *Mycolicibacterium smegmatis*). 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 [Fig. 3; AFZ14666.1, Crinalium].

## Operons with chaperones

### *PspA with CesT/Tir-like domain*

A protein with a **CesT/Tir-like domain**, a chaperone, occurs in an operon with PspA in bacteroidetes, gemmatimonadetes, alpha- and deltaproteobacteria, spirochaetes, and chlorobi [**Fig. 3**; **Table S1**; ANH61663.1, Dokdonia]. We found that another proteobacterial protein (*yjfI*, *E. coli*), identified as a DUF2170-containing protein by Pfam, was also a CesT/Tir-like domain, co-occurs with the *yjfJ* family of PspA [**Fig. 3**; **Table S1**; BAB38581.1, *E. coli*; [32]].

### *PspA with TM-Flotillin dyad and Band_7*

Further, in bacteroidetes, proteobacteria, firmicutes, and acidobacteria, PspA occurs with another chaperone (Band_7) in the form of a conserved dyad, the **TM-Flotillin dyad**, with a TM protein (Signal and 2 TM and a C-terminal domain, Pfam: DUF1449), a protein with an anchoring TM, Band_7 domain, coiled-coil linker, and the flotillin domain [**Fig. 3**; **Table S1**; ANH61663.1, Dokdonia]. In most bacteroidetes and a few proteobacteria, both CesT/Tir and TM-Flotillin dyads occur together with PspA [**Fig. 3**; **Table S1**; AAN56746.1, Shewanella]. In most of these combinations, an additional protein with beta-propeller and AAA-ATPase is seen in the operon.

We found that a few of the Flotillin operon dyads without proximal PspA have the beta-propeller/ATPase in the operon in some actinobacteria, but they all have PspA elsewhere in the genome. An example is the TM protein of the Flotillin operon dyad, belonging to the family typified by ***yuaF*** in *Bacillus subtilis*, which is implicated in the resistance to cell wall antibiotics [86]; also similar to the ***yqiJ*** family in *Escherichia coli*. Analysis of this yqiJ protein showed that the TM-Flotillin operon dyad is widespread in archaea and bacteria. Even though most of these dyads have PspA elsewhere in the genome and in the same operon, yqiK (Band_7/Flotillin containing protein with two N-terminal TM and operonic partner of yqiJ) has been shown to support the stress control function of PspA [87]. The cardiolipin-rich inner membrane and the Flotillin complex support the PspA components – either the associated members in the operon or the PspFABC operon elsewhere in the genome – to control membrane dynamics [REF].

*PspA with Spermine synthase*

PspA occurs with a membrane-associated **Spermine synthase**, a TM protein, and DUF4178 (an all beta-domain-containing protein) in gammaproteobacteria (AAN56746.1, Shewanella), firmicutes (CCG46123.1, Halobacillus), bacteroidetes, and a few alphaproteobacteria and aquificae [**Fig. 3**; **Table S1**]. In these gammaproteobacteria and bacteroidetes, the operon has the **CesT_Tir chaperone–PspA** dyad. The TM protein in this Spermine synthase containing operon is an *E. coli yjfL* (DUF350)-like TM that occurs in alpha- and gammaproteobacteria with the yjfIJKLM operon where *yjfK* and *yjfM* are uncharacterized DUF2491 and DUF1190 containing proteins respectively and yjfL-yjfJ are the CesT_Tir-DUF2170 and PspA (BAB38581.1, *E. coli*) dyad [**Fig. 3**; **Table S1**].

In some firmicutes, PspA (AAS43862.1, Bacillus) occurs with the yjFL-likeTM protein (DUF350*)* and DUF4178. In actinobacteria, the membrane-associated Spermine Synthase (Rv2601 *M. tuberculosis H37Rv*) occurs in an operon, without PspA, with the yjFL-like TM and DUF4178 protein (Rv2597, Mycobacterium) and two other uncharacterized domains. Most of these actinobacteria have a PspA elsewhere in the genome. In beta- and deltaproteobacteria, many of which do not have PspA in the genome, the membrane-associated Spermine synthase–yjfL-TM–DUF4178 triad occurs without PspA and with the amino oxidase and adenosylmethionine decarboxylase proteins and a Band_7 chaperone with a C-terminal SHOCT (in betaproteobacteria). Spermidine is a polyamine that is involved in the mechanical stability of membranes, and Spermidine synthase, speE, is implicated in its biosynthesis [88,89], the amino oxidase and decarboxylase respectively lengthening and shortening the polyamine. Taken together, the membrane-associated Spermine synthase operon regulates the spermidine synthesis and PspA ports into this operon to control the membrane thickness.

*PspA with TPM_phosphatase*

Another novel operon with a chaperone is found in some firmicutes typified by *ydjF* (ANX09535.1, Bacillus), where PspA is found with a tail-anchored protein (*ydjG*) comprising two ZnR and a globular domain, a membrane-associated **TPM_phosphatase** domain (*ydjH*) [90], and a Band_7 with a C-terminal ZnR (*ydjL*) [**Fig. 3**; **Table S1**]. The ZnR domain in this system is likely involved in membrane binding. A variation occurs in actinobacteria, wherein PspA (AKL67689.1, Streptomyces) is seen in an operon with a protein containing extracellular.

*Membrane dynamics with Chaperone*

The strong association of PspA with the CesT-Tir type chaperone of two divergent families and with the Band_7 chaperone suggests that the chaperones help in assembling the complexes/ superstructures with PspA and/or Beta-propeller, with the AAA-ATPase used for energy, whenever present. This superstructure building can be triggered at times of stress and can be used to stabilize the membrane. A similar theme is followed in the association with the Band_7 chaperone, where the Flotillin complex may be used as a base to build the PspA complex to strengthen the membrane. We found that an additional element in the form of the Beta-propeller is introduced in the complex in some bacteria. The PspA/CesT_Tir dyad and PspA by itself in some cases is recruited by the Spermine synthase operon to control the membrane stability by increasing or decreasing spermine synthesis in response to environmental stimuli [88].

### 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 (Toast_rack), LiaF (LiaI-LiaF-TM and Toast_rack), 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 S1**; *add Mascher ref;* 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 S1**; *e.g., APB74393.1, part of eight gene context: APB74392.1 to APB74398.1*]. In the actinobacterial genus, Streptomyces [**Fig. 3**; **Table S1**; CAB51252.1], the dyad of PspA and PspAA occur with a two-component system of the firmicute-type.

#### *Classic Two-component Transcriptional signaling system*

The two-component signaling system utilizes 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 response regulator REC/HTH containing protein, which regulates the expression of target genes. We note that PspA, LiaI-LiaF-TM, and Toast_rack tie into this system to strengthen the membrane in response to the stress signal. In actinobacteria where the Histidine Kinase is fused to PspC, the signal is sensed by PspC, and Toast_rack can act as the internal sensor. Moreover, other PspA systems elsewhere in the genome could be recruited in addition whenever they are not found in the immediate operonic neighborhood.

**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 and their possible histories based on similar/dissimilar domain architectures and genomic contexts. 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 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 get a more holistic understanding of the stress response system.

## PspA-free variations of domain architectures

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 phage-shock protein stress response systems. Below we have described some of our major findings related to the *Toast_rack* domain and its homologs and the PspC domain before outlining a range of novel Psp-A free genomic contexts.

## Toast_rack domain and the N-terminal transmembrane region LiaI-LiaF-TM

### *Toast_rack Domain Architectures*

We first examined the **Toast_rack** domain and all its protein homologs obtained from various starting points because we found it to be the most widely-spread domain [**Fig. 4**; **Table S2**; see Methods]. The first notable finding is that the Toast_rack domains in most proteins are C-terminal with an N-terminal transmembrane region comprising the defined PspC, LiaI-LiaF-TM, other generic TMs, or a combination thereof [**Fig. 4**]. We also found versions with an N-terminal anchoring TM region or with N-terminal alphahelical domains. Analysis of the Toast_rack containing proteins revealed various fusions including two previously uncharacterized N-terminal alphahelical domains described below.

### *Toast_rack in actinobacteria and firmicutes: SHOCT and SHOCT-like proteins*

Toast_rack containing proteins in actinobacteria (CCP43715.1, Mycobacterium) with a few transfers to proteobacteria, spirochaetes, and other bacteria are fused to an N-terminal **SHOCT-bihelical domain** (Pfam: **DUF1707** [91]) and a C-terminal Toast_rack [**Fig. 4**]. Based on sequence analysis, we predict that the SHOCT domain will act as an anchoring domain to the sub-membrane region.

We found another novel variant of the SHOCT domain, with a bihelical secondary structure (that we call the **SHOCT-like** domain to distinguish from the SHOCT domain), fused to the N-terminus of Toast_rack domain-containing protein typified by *Bacillus subtilis* ***yvlB*** (**SHOCT-like** is at the C-terminus of the current **DUF2089** profile in Pfam. CAB15517.1, Bacillus) [**Fig. 4**; **Table S2**]. YvlB has been implicated in the activation of the LiaFSR operon [92]. In a few members of the Lactobacillus and Enterococcus genera, a coiled-coil linker is found between the two domains (SHOCT-like and Toast_rack; AQL53774.1) [**Fig. 4**; **Table S2**]. Analysis of the **SHOCT-like** domain showed that it has independent and widespread existence with a different C-terminal globular domain (BMQ_3691, Bacillus) [**Fig. 4**; **Table S2**]. Several of the members of the BMQ_3691 family occur in an operon with another divergent version of the domain (BMQ_3692) with an N-terminal ZnR fusion, a DUF2089-HTH in the middle, and a C-terminal SHOCT-like domain. Both the protein families BMQ_3691/3692 are found in firmicutes, chloroflexi, thermotogae, deinococcus, armatimonadetes, a few actinobacteria, and unclassified bacteria [**Fig. 4**; **Table S2**].

### *Toast_rack in Bacteroidetes: DUF1700 and DUF1700-like proteins*

We find another recurring transmembrane fusion theme in Toast_rack containing proteins in Bacteroidetes (AOW09537.1, Flavobacterium), firmicutes, betaproteobacteria, and a few other bacterial clades, with an **N-terminal DUF1700-like alphahelical domain** (that we call, DUF1700-ahelical), followed by PspC, LiaI-LiaF-TM or other TM regions and a C-terminal Toast_rack domain (Pfam: DUF1700, DUF1129, clan Yip1) [**Fig. 4**; **Table S2**]. The DUF1700-like alpha-helical domain belongs to the Yip1 clan in PFAM. An HHpred search with the alignment retrieves PDB structure 2O3LA of DUF1048 family, mostly in firmicutes, belonging to the same clan. These members also have PadR-HTH in the operon. Analysis of the sequence alignment shows that the structure has three core helices, and the third helix has a bump in the middle due to a conserved GxP motif, which forms the edge of a binding groove. Other than its fusions to Toast_rack, the DUF1700-like alphahelical domain is fused to either intracellular sensor domain like the Pentapeptide repeats [AOH56696.1, Bacillus] (mostly in firmicutes), or to various transmembrane or intramembrane sensor domains, instead of PspC and LiaI-LiaF-TM [**Fig. 4**; **Table S1**]. The DUF1700-ahelical domain also occurs at the N-terminus of distinct TM proteins such as the MacB_PCD extracellular domain

(PDB:3IS6_B -like) containing FtsX proteins (acidobacteria, gemmatimonadetes, verrucomicrobia with transfers to few other clades), certain members of the firmicutes FtsW/RodA/SpoVE cell cycle family proteins [**Fig. 4**; **Table S1**; CAC98500.1, Listeria], the vanZ protein (actinobacteria), and various other 6-TM, 4-TM and 2-TM proteins. While the 4-TM family forms a major group and is likely to be a divergent LiaI-LiaF-TM domain, members of the 6-TM family [**Fig. 4**; **Table S1**; AAO82145.1, Enterococcus] are fused at the N-terminus either to a wHTH or to the DUF1700-like alphahelical domain, while some paralogs are fused sporadically to the SHOCT-like bihelical domain. The N-terminal TMs or alphahelical domains help attach Toast_rack to the membrane or sub-membrane where it acts as a sensor domain.

### *Toast_rack in Cyanobacteria*

Toast_rack also has sporadic fusions to **Caspase** in cyanobacteria [**Fig. 4**; **Table S2**; AFY83227.1, Oscillatoria]. Our analysis of the topology of the above architectures indicates that Toast_rack is an intracellular domain.

### *Toast_rack with LiaI-LiaF-TM or PspC*

**LiaI-LiaF-TM** mainly occurs in fusion with Toast_rack (CAB15300.1, Bacillus) or **PspC** (ABY34522.1, Chloroflexus) or a combination of both [**Fig. 4**; **Table S2**; ACV77657.1, Nakamurella]. Another fusion is with an N-terminal PspC and C-terminal Histidine Kinase found in actinobacteria (sometimes in the neighborhood of Toast_rack; AIJ13865.1, Streptomyces) [**Fig. 4**; **Table S2**]. Sporadic fusions are also found to N-terminal BBOX or the SHOCT-like domain.

### PspC domain architectures

PspC by itself is present in most bacteria and in the archaeal clades of euryarchaeota and Asgard. Some orthologs of the solo domain occur with an N-terminal Zinc Ribbon fusion sporadically in various clades [**Fig. 4**; ABC83427.1, Anaeromyxobacter]. Almost all other PspC homologs show fusions with one or more transmembrane domains [**Fig. 4**; **Table S2**]. The various fusions with Toast_rack, LiaI-LiaF-TM, DUF1700-ahelical domain, and with Histidine Kinase have been already discussed above. The PspC protein of gammaproteobacteria (in PspFABC) carries a fusion to a C-terminal PspB with sporadic transfer to spirochaetes, nitrospirae and unclassified bacteria.

**In summary**, we have discovered several novel domain architectures containing Psp partner domains, in addition to those previously reported. For instance, we have identified several lineage-specific domain architectures of Toast_rack with N-terminal transmembrane domain fusions, such as DUF1700-ahelical in bacteroidetes, LiaI-LiaF-TM in firmicutes, SHOCT-bihelical and SHOCT-like domains in actinobacteria and firmicutes, and PspC in actinobacteria [**Fig. 4**, **Table S2**]. We also found that, in addition to solo PspC domains being the most prevalent, PspC can be fused to PspB, DUF1700-ahelical, Toast_rack, and LiaI-LiaF-TM domains [**Fig. 4**, **Table S2**]. All variations of domain architectures that do not contain PspA can be explored in the webapp.

## Novel PspA-free Genomic contexts

Since genomic contexts are critical to the function of bacterial and archaeal proteins, we next explored the neighborhoods of each of the 'PspA-free' homologs of the Psp partner domains.

## Operons of LiaI-LiaF-TM, Toast_rack, and PspC with two-component systems

In addition to the Lia operon discussed above in the section on '*PspA: Operons with two-component systems*,' we found that a few orthologs of LiaF, typified by *vraT* of *Staphylococcus aureus* [ABD31150.1; **Fig. 5**; **Table S2**] and other firmicutes, that are part of PspA-free contexts. These orthologs have LiaI-LiaF-TM and Toast_rack with the two-component system (*vraSR*) and are involved in methicillin and cell-wall antibiotic resistance [5,93–95]. A similar organization of LiaI-LiaF-TM and Toast_rack containing protein with a two-component system is found in bacteroidetes, a few acidobacteria, beta- and gammaproteobacteria (ABD83157.1), and Ignavibacteriae (AFH48155.1) [**Fig. 5**; **Table S2**].

Unlike the PspA-associated Lia-like two-component systems (detailed in the PspA genomic context section above), in most other actinobacteria, a different type of Histidine Kinase, one with an N-terminal PspC, followed by LiaF-LiaI-TM region and a terminal Histidine Kinase (*e.g.,* BAB98004.1, Corynebacterium) occurs with a Receiver/HTH-containing gene [**Fig. 5**; **Table S2**]. PspA is not seen in these contexts. This dyad shares a promoter region with a protein containing PspC, additional TMs, and Toast_rack [**Fig. 5**; **Table S2**]. Some actinobacterial genomes (*e.g.,* BAD60342.1, Nocardia) have an additional two-component dyad of the firmicute type with only a SHOCT-bihelical and Toast_rack containing protein in the operon. In addition, in a few deinococci, a Toast_rack protein (ACO46500.1) is sandwiched between a two-component dyad [**Fig. 5**; **Table S2**].

## Operons with transcriptional regulators

### *DUF1700 with PadR-HTH*

The DUF1700-like alphahelical domain (HHpred 2O3L_A) in all its fusions almost always occurs in a tightly associated operon with a wHTH (Pfam: **PadR**; *lltR*/*lstR* in *Listeria monocytogenes* [96]]). Given the widespread phyletic spread of the domain and its association with this operon, the operon can be traced back to the base of the bacterial lineage. In the Psp system, the DUF1700 domain is either found fused to PspC, LiaI-LiaF-TM, and Toast_rack or occurs in an operon with them [**Fig. 5**; **Table S2**; AKD55949.1, Spirosoma]. In all these cases the PadR-like wHTH is found in the operon. In the family with 6TM domain (AAO82145.1, Enterococcus), that is fused to either a wHTH at the N-terminus or to the DUF1700-ahelical domain, the PadR-HTH is found in the operon only in fusions to the DUF1700-ahelical domain [**Fig. 5**; **Table S2**].

### *DUF1700 with wHTH*

DUF1700-ahelical with TM (typified by CAC98678.1, Listeria) is flanked by *lltR* encoding a wHTH domain, and lmo0601 encoding a Toast_rack protein [96] [**Fig. 5**; **Table S2**]. DUF1700-ahelical fused to FTSW_RODA_SPOVE typified by lmo0421 (CAC98500.1, Listeria) has lmo0422/LstR (*Listeria*) encoding a PadR-wHTH, and an additional gene lmo0423 encoding a Sigma-HTH in the operon [96,97] [**Fig. 5**; **Table S2**]. Fusions to MacB_PCD/ FtsX also have the PadR-HTH-like wHTH in the operon, while members of the family without the fusion do not have the wHTH in the operon [**Fig. 5**; **Table S2**].

The only instance where wHTH occurs without the DUF1700 alphahelical domain-containing Toast_rack is in a family of intracellular Toast_rack in actinobacteria typified by (CAC05954.1) in

*Streptomyces coelicolor*. The DNA-binding domain **LytTR** is found in an operon with Toast_rack fused to LiaI-LiaF-TM (LA14_0400) in firmicutes Clostridium and Lactobacillus [**Fig. 5**; **Table S2**].

*DUF2089-HTH*

The **DUF2089-HTH** containing BMQ_3692 family, with N-terminal ZnR fusion and C-terminal SHOCT-like domain is found in an operon with the SHOCT-like and the globular C-terminal domain (BMQ_3691, *Bacillus megaterium*) in firmicutes, chloroflexi, thermotogae, deinococcus, armatimonadetes, unclassified bacteria, and a few actinobacteria [**Fig. 5**; **Table S2**]. Some members (*e.g.,* DR_0017, *Deinococcus radiodurans*) have a Toast_rack with N-terminal SHOCT-like (AAF09613.1, Deinococcus) in the operon in addition to or instead of the SHOCT-like and the globular C-terminal domain (BMQ_3691, *Bacillus megaterium*) [**Fig. 5**; **Table S2**]. The versions of the BMQ_3692 family without the N-terminal ZnR (*e.g.,* ADJ59725.1 LLNZ_03680 of *Lactococcus lactis*) are found in an operon with a 2 TM protein [**Fig. 5**; **Table S2**]. The versions of the BMQ_3692 family without the C-terminal SHOCT-like domain, but retaining an intact ZnR, in Bacteroidetes is found in an operon with a secreted AB-Hydrolase, a membrane protein, and a membrane attached PH domain (*e.g.,* AIL47037.1, Elizabethkingia) [**Fig. 5**; **Table S2**].

*SIGMA-HTH*

We found that various Bacteroidetes, firmicutes, acidobacteria, verrucomicrobia, gemmatimonadetes, few chloroflexi, and actinobacteria contain a triad of **SIGMA-HTH** followed by a protein with ZF and TPR connected by a Transmembrane Helix (with some lacking the ZF) and an anchored Toast_rack domain [**Fig. 5**; **Table S2**]. The third gene in this system may be followed by LiaI-LiaF-TM and Toast_rack combinations. In a few acidobacteria (AEU34960.1, Granulicella), a protein with BBOX domain fused to a LiaI-LiaF-TM is found in the operon [**Fig. 5**; **Table S2**].

*GNTR-HTH and TraJ-RHH*

A transport operon with ABC-ATPase and a transporter protein along with a protein with **GNTR-HTH** and a C-terminal alphahelical domain and a protein with an anchored Toast_rack (typified by AAM36414.1, Xanthomonas) in the operon is found in gammaproteobacteria and a few Bacteroidetes, firmicutes, and spirochaetes [**Fig. 5**; **Table S2**]. Many of these operons have a DUF2884 domain-containing lipoprotein. A convergent operon is found in actinobacteria where a Toast_rack containing protein occurs with the transport operon along with a TraJ-RHH containing protein instead of the GNTR-HTH, while B737_2034 in *Amycolatopsis mediterranei* (AGT82698.1) has an additional GNTR-HTH fused to an FCD small molecule binding domain in the operon. The TraJ-RHH containing protein has an N-terminal alphahelical domain and a C-terminal strand.

*SHOCT-like domains*

The **SHOCT-like** fusion with Toast_rack (*yvlB/ythC*) occurs in an operon with PspC (*ythA/ythB/YvlC*) and other membrane proteins (*yvlA* and *yvlD*) in firmicutes like yvlB of *Bacillus subtilis subsp. natto BEST195* (BAI87136.2) [**Fig. 5**; **Table S2**]. YvlB has been implicated in the activation of the LiaFSR operon [92]. The ythA/ythB/YvlC family of PspC has been shown to contribute to protection against membrane permeabilization [59,98].

The transport operon with ABC-ATPase and a transporter protein is found in an operon with an N-terminally anchored Toast_rack containing protein sporadically in firmicutes (AJI31452.1 *Bacillus*

*cereus* BF28_3762), actinobacteria (two tandem proteins with Toast_rack ALR80200.1 *Streptococcus salivarius*) and euryarchaeota [**Fig. 5**; **Table S2**]. Some actinobacterial DUF1707-SHOCT-like bihelical domain and Toast_rack containing protein also occur with the ABC transport operon with two permease proteins (CAB88834.1 *Streptomyces coelicolor A3* SCO2893) [**Fig. 5**; **Table S2**]. Some of the aforementioned ABC transporter operons have a two-component system inserted in the operon.

*Lessons from the Novel Transcriptional Signaling systems*

The obligate nature of the pairing of the DUF1700-ahelical domain and PadR-HTH suggests that they bind to each other and, upon sensing a stimulus by the PspC or equivalent TM region fused to the DUF1700-ahelical domain, the PadR-HTH is released to induce transcription. This kind of signal transmission is similar to the two-component system, where the signal is transferred by phosphorylation of the REC domain. Thus, the DUF1700-ahelical domain and PadR-HTH define a novel signaling paradigm, and PspC and Toast_rack elements of the Psp system are ported into this novel signaling system, whereby it regulates the membrane dynamics. In the cases where Toast_rack and LytTR-HTH occur in an operon together or as a fusion without the DUF1700-ahelical domain, Toast_rack is predicted to perform a similar function.

The DUF2089-HTH, Toast_rack, and SHOCT-like systems define a novel transcriptional regulatory system where the SHOCT-like domain is used to attach to the membrane, and the signal is transmitted by Toast_rack. The GNTR-HTH and TraJ-RHH associated operons port the Toast_rack into an ABC transporter system, which allows control of transport and links it to membrane strengthening. While these operons do not have PspA or PspC, the organisms have these psp elements elsewhere in the genome, and these would be recruited into the system as demonstrated by the PadR-HTH deletion experiments affecting the Lia system in Listeria [96].

**In summary**, in addition to novel extensions and variations of the previously known Psp operons (with PspBC, LiaI-LiaF-TM/Toast_rack, and two-component systems), we have discovered and discussed in detail several novel players and related genomic contexts. For example, we found several novel transcriptional signaling systems (such as PadR-HTH, wHTH, DUF2089-HTH, SIGMA-HTH, GNTR-HTH, and TraJ-RHH) co-occurring with Toast_rack, DUF1700-ahelical, and SHOCT-like proteins [**Fig. 5**, **Table S2**]. Our study has identified several new extensions to the membrane stress response system. To conclude our study, we finally proceeded to consolidate all our findings in the context of PspA and Psp partner domains to identify the extended network of stress response players in bacteria, archaea, and eukaryotes.

## Psp Consolidated

In the sections above, we have identified a variety of 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 Toast_rack, the most wide-spread PspA partners, as well as the domains that frequently co-occur with our initial set of proteins of interest. We next asked how to reconcile the different pieces of the Psp puzzle that we have identified.

### The proximity network of PspA and cognate partner domains

To generate a global picture, we generated a network representation that is based on the domain architectures of the larger Psp system reported above [**Fig. 6A**]. The domains are the nodes of this

network, and they are connected pair-wise if they co-occur in a single protein. The network thus summarizes the connections between PspA, Psp cognate partner domains (such as PspC and Toast_rack), and other novel domain connections such as DUF1700-ahelical and SHOCT-bihelical/SHOCT-like proteins.

### The phyletic spread of Psp members and their prevalent partner domains

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

### The prevalence of most common neighbors of Psp members

We next adopted the upset plot to summarize the overlap of several sets of domain architectures concurrently. Using this visualization, we quantified the relative occurrences 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 found that the Toast_rack and LiaI-LiaF-TM context that contains the two-component system and DUF1700-ahelical co-occurring with PadR-HTH are among the most prevalent neighborhoods closely following singletons such as PspA and PspC without characterized neighborhoods. It is noted that Snf7 appearing alone most frequently is an artifact of the genomic context analysis being restricted (and relevant) to bacteria and archaea.

**In summary**, these visualizations help us build 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**].

## Conclusions

Recent work with the phage shock protein (Psp) stress response system in bacteria suggests that even though Psp function appears to be maintained across phyla, the sequence and structure of the accessory Psp proteins, the regulatory mechanisms, and the membrane stress response dynamics of the system vary widely among bacterial species [12,14,38,61]. PspA has also been shown to be needed for maintaining the cell shape in Corynebacterium [85]. Studies in archaea show that PspA is involved in cell division [35]. In eukaryotes, PspA is involved in thylakoid biogenesis and vesicular budding [99]. These variations likely reflect different envelope structures and genetic and environmental factors that necessitate the concomitant evolution of these membrane stress response systems.

In this study, we report the results of a comprehensive and detailed analysis of the evolution of the Psp system and all its partner proteins, focusing on the phylogenetic distribution, sequence/structural features of PspA, its known and unknown partners, and their genomic neighborhoods across the tree of life. We used this analysis to explore and identify novel

components in the Psp stress response system. Notably, we first establish that PspA is indeed universal and that the ancestry of the PspA/Snf7 superfamily (with the eukaryotic ESCRT system) dates back to the last universal common ancestor (LUCA). We also define novel domains such as LiaI-LiaF-TM and Toast_rack. We then go beyond the global phyletic distribution of PspA and its known partners, to identify distinct and commonly occurring themes (domain architectures and genomic contexts) in Psp systems that are both widely distributed (*e.g.,* PspA/Snf7, PspC, and Toast_rack) and restricted to specific lineages (*e.g.,* discovering DUF1700-ahelical and SHOCT-bihelical often in the proximity of Toast_rack, and mostly in firmicutes and actinobacteria, respectively). We also find several novel proximal domains including PspA/Snf7 alongside AAA-ATPase, PspA with PspAA (a novel PspA-associated domain that we find in actinobacteria and archaea), and PspC often co-occurring with Toast_rack and two-component systems in actinobacteria rather the 'classic' proteobacterial context. Finally, in addition to novel extensions and variations of known Psp operons (with PspBC, LiaI-LiaF-TM/Toast_rack, and two-component systems), we have discovered several novel components such new transcriptional signaling systems and their pertinent genomic contexts.

We present all the findings (data, results, and visual summaries) from this study for in-depth exploration 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 (~6500 completed representative/reference genomes with taxonomic lineages; NCBI NR database [100,101]; All Psp homologs are listed in the webapp; representatives in **Table S1, S2**). The phyletic order (sequence) was obtained from NCBI taxonomy and PATRIC [100–105].

## 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 [106] 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 of genomes carrying each of the query proteins were recorded [webapp]. These proteins were clustered into orthologous families using the similarity-based clustering program BLASTCLUST [107]. HHPred, SignalP, TMHMM, Phobius, JPred, Pfam, and custom profile databases [108–115] were used to identify signal peptides, transmembrane regions, known domains, and the secondary protein structures in every genome. Homolog

information including domain architectures for each of the Psp member proteins is shown in the webapp ('Data' and 'Domain architectures' tabs).

## Neighborhood search

Bacterial gene neighborhoods — ± 7 genes flanking each protein homolog — were retrieved using custom scripts from GenBank [100,101]. Gene orientation, domains, and secondary structures of the neighboring proteins were characterized using the same methods applied to query homologs above. Genomic contexts for each of the Psp members are shown in the webapp ('Genomic contexts' tab).

## Phylogenetic analysis

Multiple sequence alignment (MSA) of the identified homologs was performed using Kalign [116,117] and MUSCLE [118,119]. The phylogenetic tree was constructed using FastTree 2.1 with default parameters [120]. MSA and phylogenetic trees for each of the Psp members are shown in the webapp ('Phylogeny' tab).

## Network reconstruction

TThe Psp proximal neighborhood network was constructed based on the domain architectures and genomic contexts of PspA and its cognate partner proteins. The nodes represented the domains and edges indicated a shared neighborhood (domains of the same protein or neighboring proteins). Proximity networks for each of the Psp member proteins are shown in the webapp ('Domain architectures' tab).

## Web application

The interactive and queriable web application was built using R Shiny [121,122]. All data analyses and visualizations were carried out using R/RStudio [123,124] and R-packages [125–127,127–137]. 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* (liaIHGFSR) are shown. Boxes indicate genes/proteins and colors indicate 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, Toast_rack. 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.

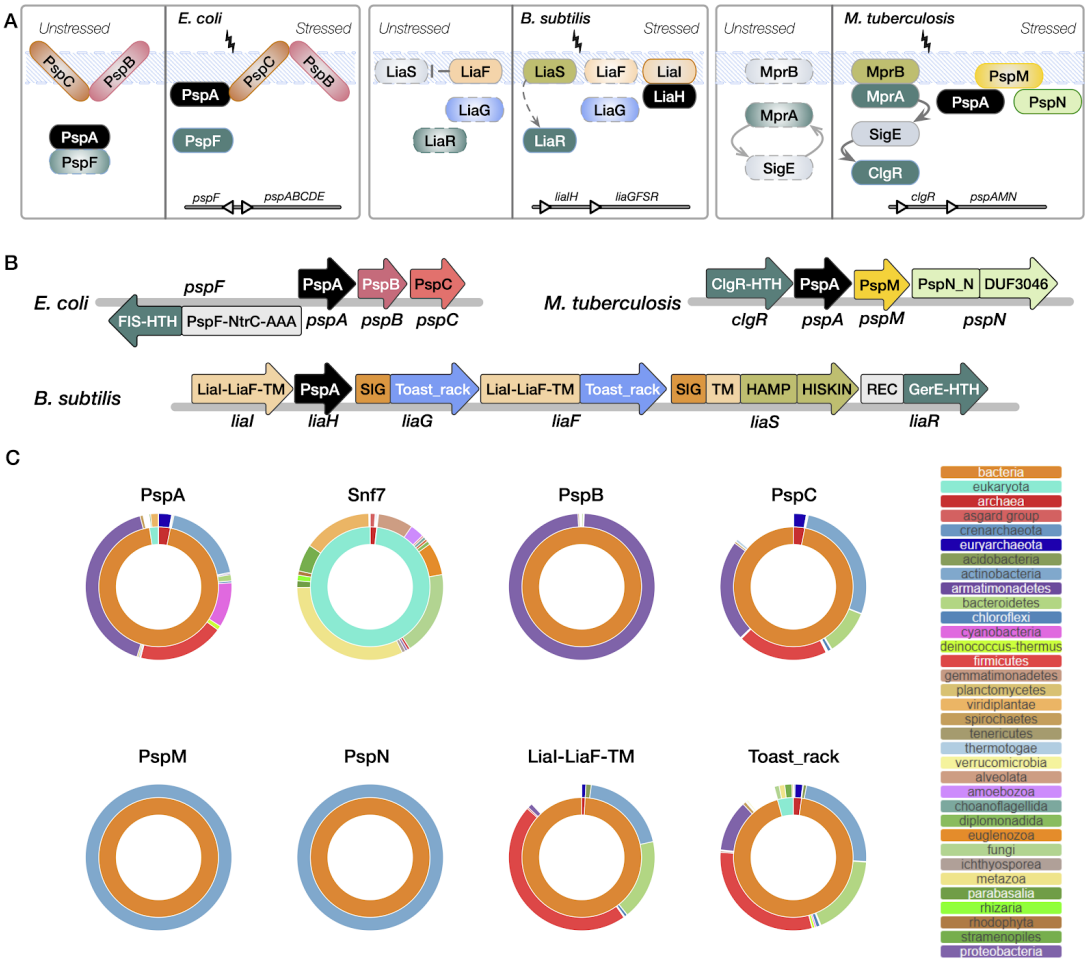


**Figure 2. Phylogenetic tree of PspA homologs across the tree of life.**

The phylogenetic tree was constructed based on a multiple sequence alignment performed using representative PspA homologs across all the major kingdoms/phyla (see Methods). 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.

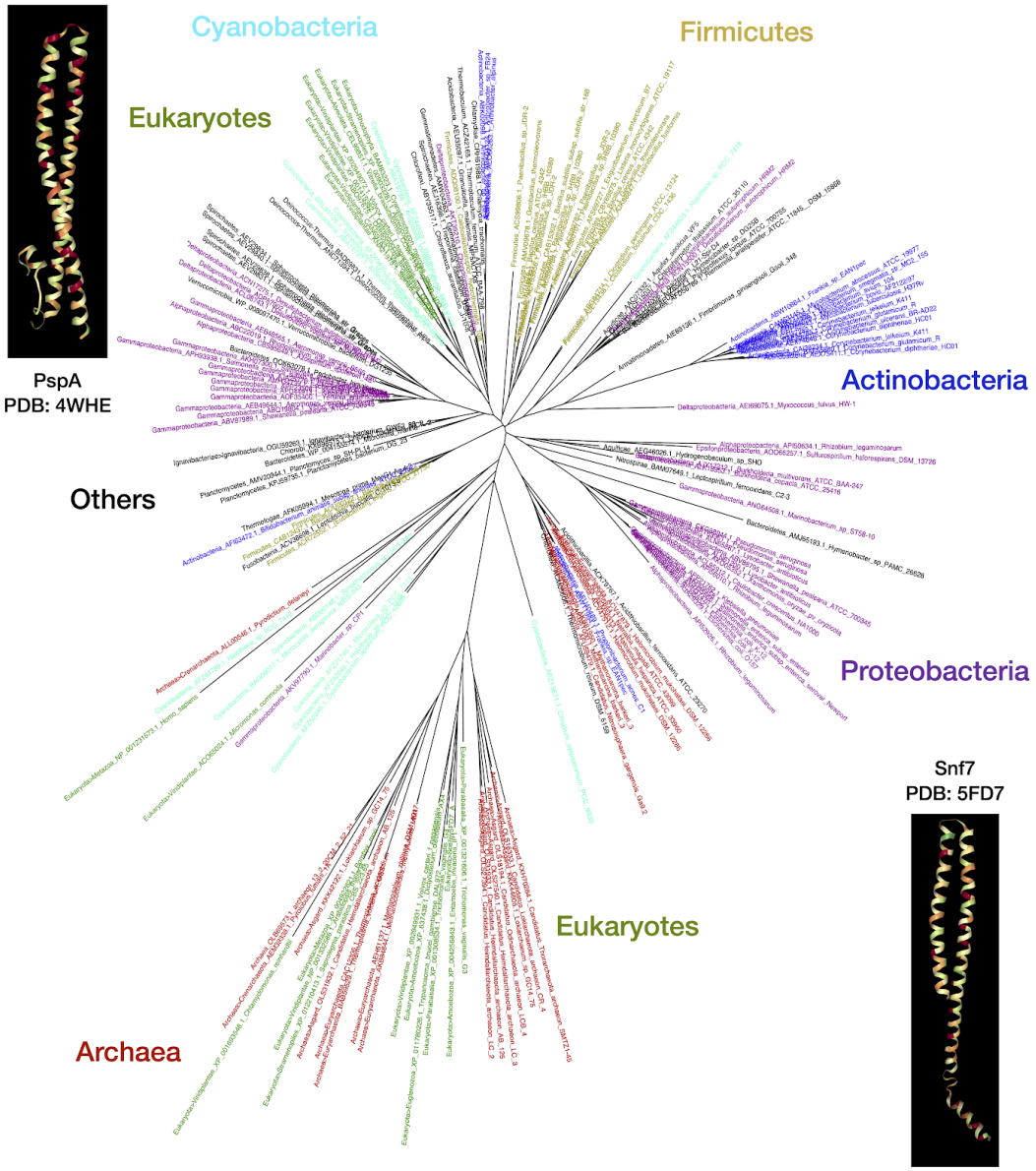


**Figure 3. PspA domain architectures and genomic contexts**

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 characteristic 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; (c) ClgR, PspM/PspN, Thioredoxin (**Fig. S1**); (d) chaperones such as Band_7, Flotillin, CesT-Tir, TPM_phosphatase, ZnR, SpermGS-ATPgrasp, and Spermine synthase; (e) two-component systems such as the Lia system and Toast_rack, and other novel genomic contexts; and (f) cyanobacterial variations. Representation of the predominant contexts were identified by neighborhood searches (±7 genes flanking each homolog; see Methods). Key: *rectangles*, domains; *arrow-heads*, 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 change in 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. Characteristic 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 S1**, 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). *Add * notes here.*

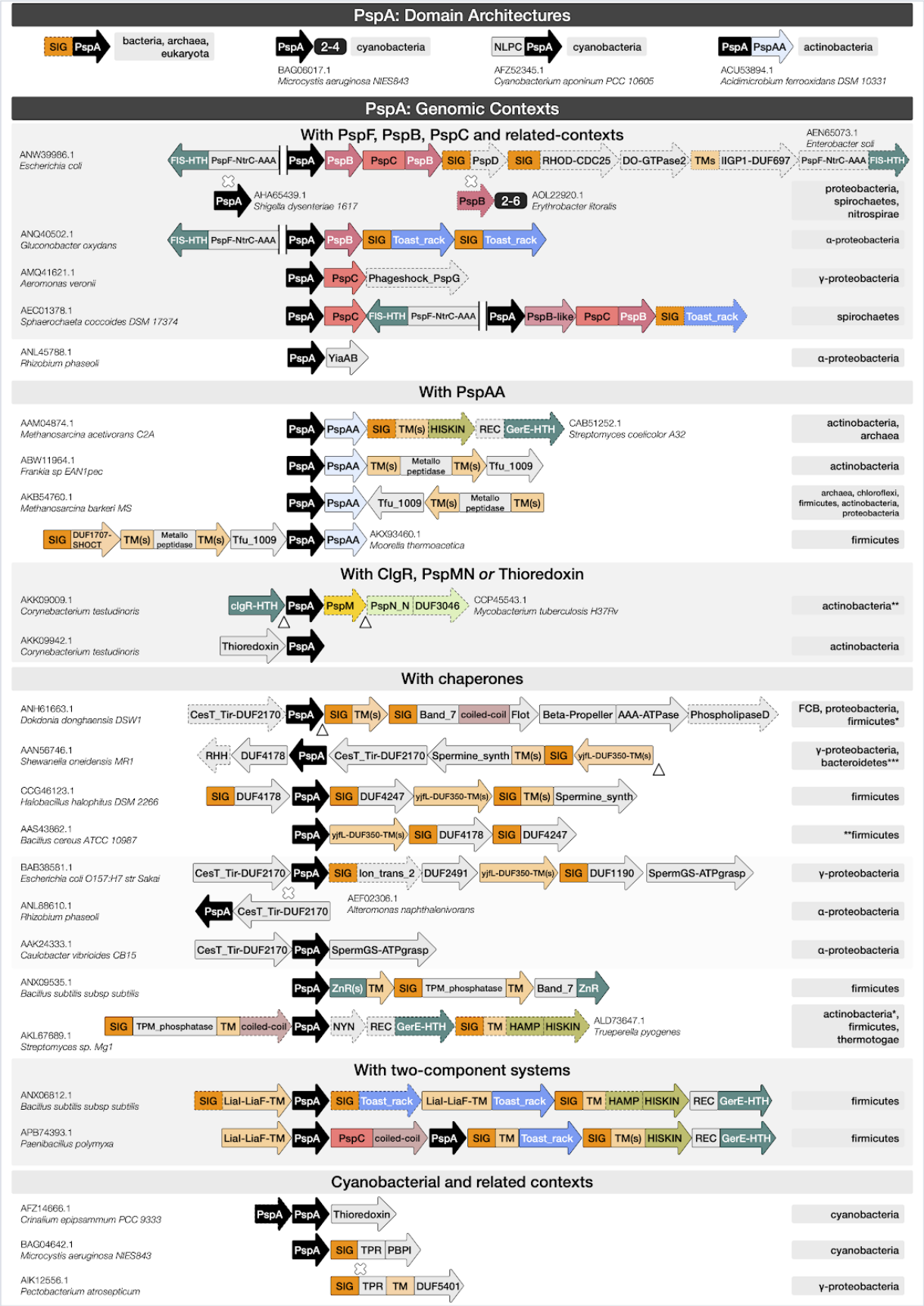

PspA: Domain Architectures
SIG PspA
bacteria, archaea, eukaryota
PspA 2-4
cyanobacteria
BAG06017.1
Microcystis aeruginosa NIES843
NLPC PspA
cyanobacteria
AFZ52345.1
Cyanobacterium aponinum PCC 10605
PspA PspAA
actinobacteria
ACU53894.1
Acidimicrobium ferrooxidans DSM 10331
PspA: Genomic Contexts
With PspF, PspB, PspC and related-contexts
ANW39986.1
Escherichia coli
FIS-HTH PspF-NtrC-AAA PspA PspB PspC PspB SIG PspD SIG RHOD-CDC25 DO-GTPase2 TMs IIGP1-DUF697 PspF-NtrC-AAA FIS-HTH
AEN65073.1
Enterobacter soli
PspA
AHA65439.1
Shigella dysenteriae 1617
PspB 2-6
AOL22920.1
Erythrobacter litoralis
proteobacteria, spirochaetes, nitrospirae
ANQ40502.1
Gluconobacter oxydans
FIS-HTH PspF-NtrC-AAA PspA PspB SIG Toast_rack SIG Toast_rack
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 Toast_rack
spirochaetes
ANL45788.1
Rhizobium phaseoli
PspA YiaAB
α-proteobacteria
With PspAA
AAM04874.1
Methanosarcina acetivorans C2A
PspA PspAA SIG TM(s) HISKIN REC GerE-HTH
CAB51252.1
Streptomyces coelicolor A32
actinobacteria, archaea
ABW11964.1
Frankia sp EAN1pec
PspA PspAA TM(s) Metallo peptidase TM(s) Tfu_1009
actinobacteria
AKB54760.1
Methanosarcina barkeri MS
PspA PspAA Tfu_1009 TM(s) Metallo peptidase TM(s)
archaea, chloroflexi, firmicutes, actinobacteria, proteobacteria
SIG DUF1707-SHOCT TM(s) Metallo peptidase TM(s) Tfu_1009 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 Beta-Propeller AAA-ATPase PhospholipaseD
FCB, proteobacteria, firmicutes*
AAN56746.1
Shewanella oneidensis MR1
RHH DUF4178 PspA CesT_Tir-DUF2170 Spermine_synth TM(s) SIG yjfL-DUF350-TM(s)
γ-proteobacteria, bacteroidetes***
CCG46123.1
Halobacillus halophilus DSM 2266
SIG DUF4178 PspA SIG DUF4247 yjfL-DUF350-TM(s) SIG TM(s) Spermine_synth
firmicutes
AAS43862.1
Bacillus cereus ATCC 10987
PspA yjfL-DUF350-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-DUF350-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 GerE-HTH 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 Toast_rack LiaI-LiaF-TM Toast_rack SIG TM HAMP HISKIN REC GerE-HTH
firmicutes
APB74393.1
Paenibacillus polymyxa
LiaI-LiaF-TM PspA PspC coiled-coil PspA SIG TM Toast_rack SIG TM(s) HISKIN REC GerE-HTH
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 including known (Toast_rack, LiaI-LiaF-TM, PspBC, PspMN, DUF3046) and other novel neighbors (DUF1700-ahelical, SHOCT-bihelical, SHOCT-like domains) are illustrated on the left. The phyletic spread of the various domain architectures are 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 S2**.

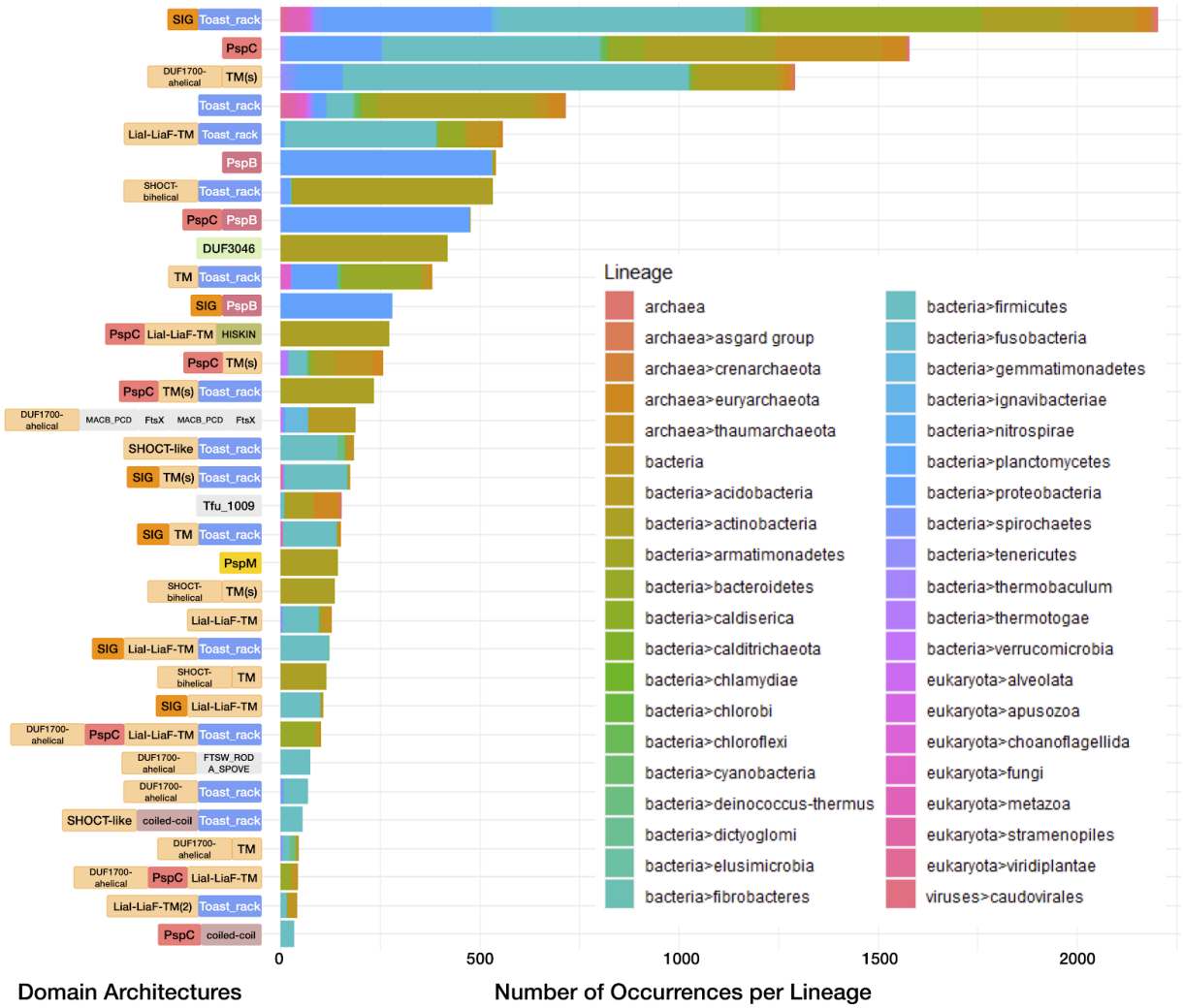


**Figure 5. The genomic contexts housing all Psp cognate partner domain homologs.**

The genomic contexts are presented using the same schematic as in **Figure 3**. The focus is on Psp partner domains such as Toast_rack (blue), PspC, LiaI-LiaF-TM, DUF1700-ahelical, SHOCT-bihelical (in shades of orange), and the various genomic neighborhoods with SHOCT-like proteins, transcription regulators (PadR-HTH, SIGMA-HTH, GerE-HTH, LytTR, GNTR-HTH), and two-component systems. 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 S2**.

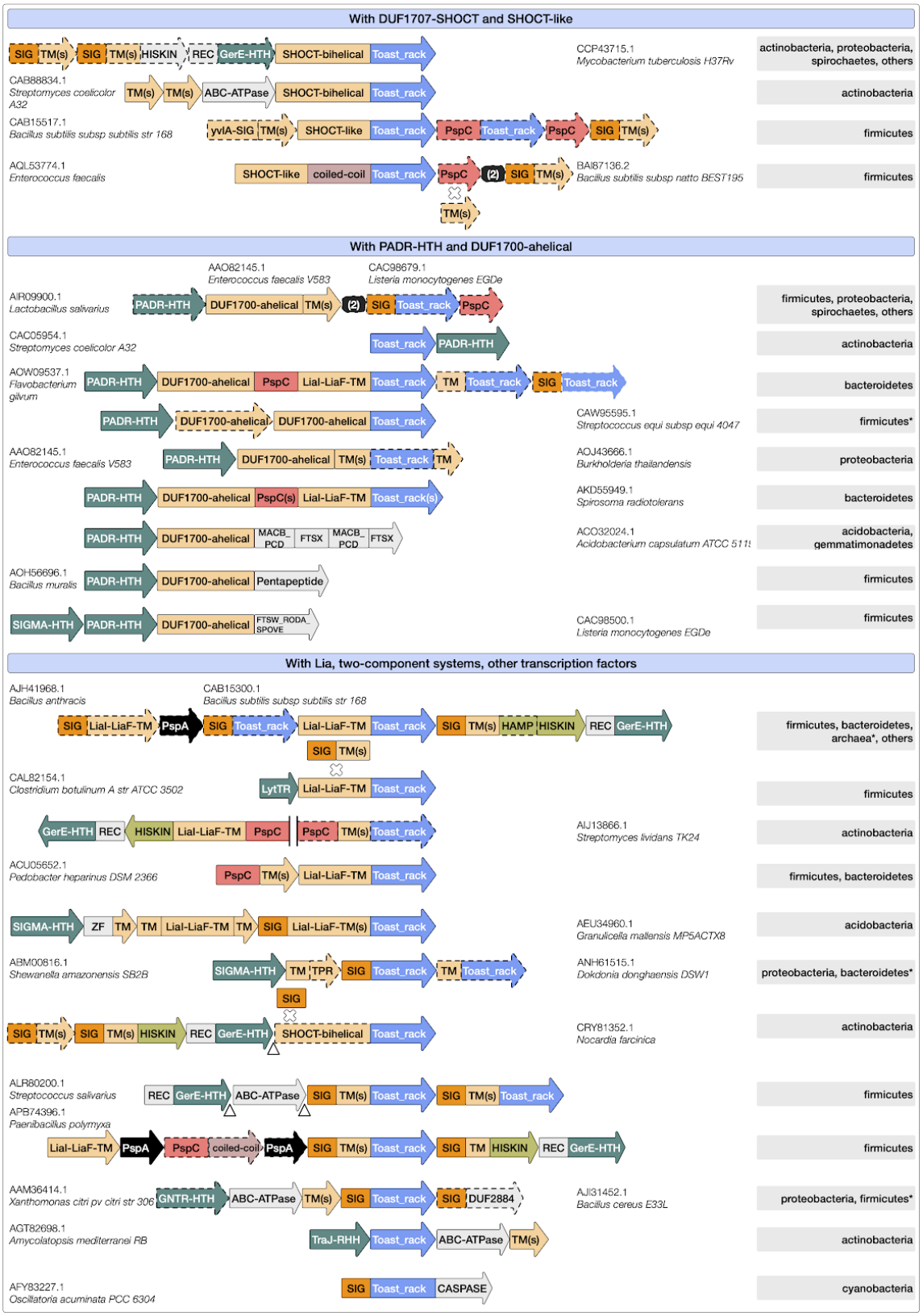
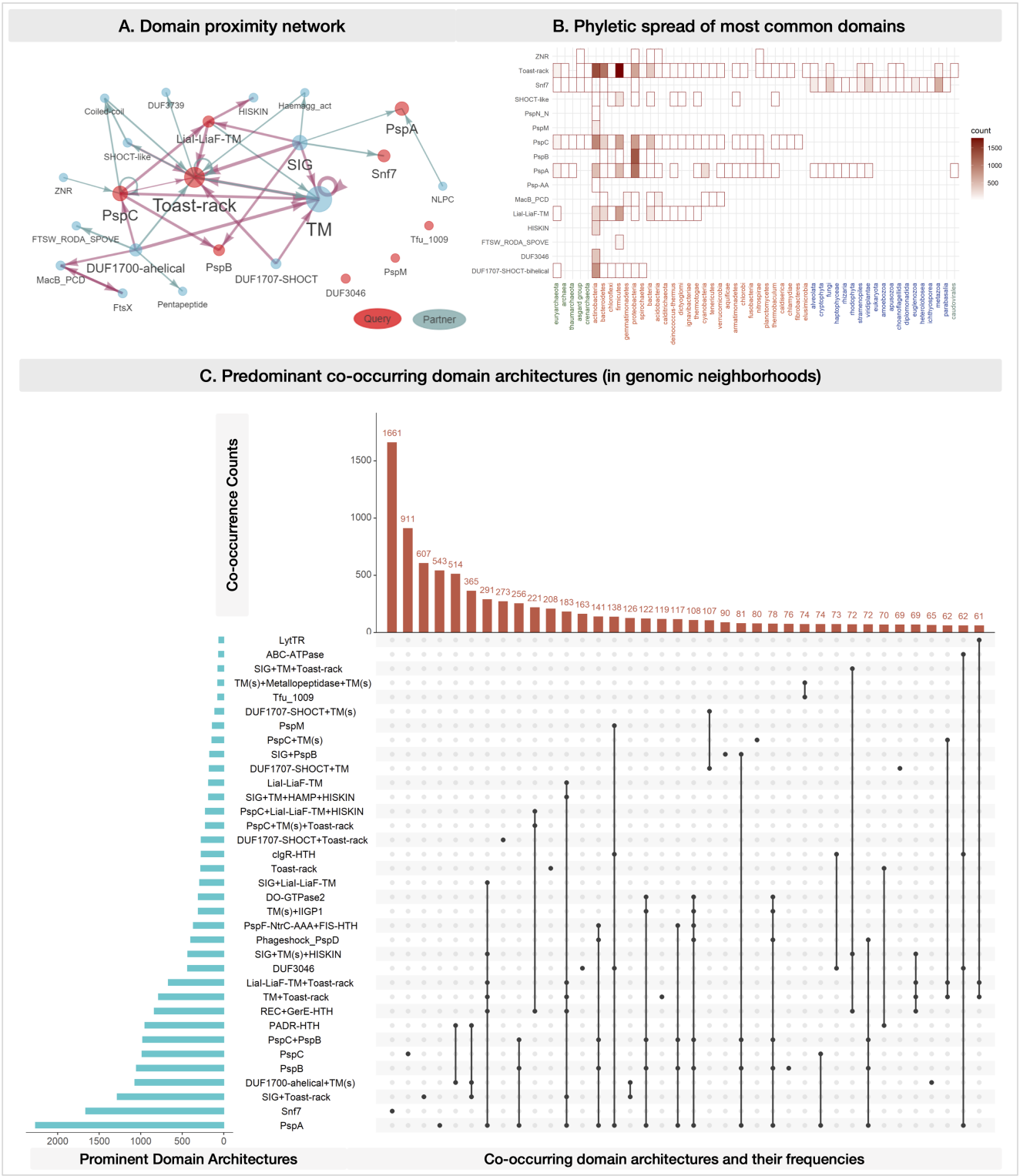

With DUF1707-SHOCT and SHOCT-like
SIG TM(s) SIG TM(s) HISKIN REC GerE-HTH SHOCT-bihelical Toast_rack
CCP43715.1
Mycobacterium tuberculosis H37Rv
actinobacteria, proteobacteria, spirochaetes, others
CAB88834.1
Streptomyces coelicolor A32
TM(s) TM(s) ABC-ATPase SHOCT-bihelical Toast_rack
actinobacteria
CAB15517.1
Bacillus subtilis subsp subtilis str 168
yvlA-SIG TM(s) SHOCT-like Toast_rack PspC Toast_rack PspC SIG TM(s)
firmicutes
AQL53774.1
Enterococcus faecalis
SHOCT-like coiled-coil Toast_rack PspC (2) SIG TM(s)
TM(s)
BAI87136.2
Bacillus subtilis subsp natto BEST195
firmicutes
With PADR-HTH and DUF1700-ahelical
AAO82145.1
Enterococcus faecalis V583
CAC98679.1
Listeria monocytogenes EGDe
AIR09900.1
Lactobacillus salivarius
PADR-HTH DUF1700-ahelical TM(s) (2) SIG Toast_rack PspC
firmicutes, proteobacteria, spirochaetes, others
CAC05954.1
Streptomyces coelicolor A32
Toast_rack PADR-HTH
actinobacteria
AOW09537.1
Flavobacterium gilvum
PADR-HTH DUF1700-ahelical PspC LiaI-LiaF-TM Toast_rack TM Toast_rack SIG Toast_rack
bacteroidetes
PADR-HTH DUF1700-ahelical DUF1700-ahelical Toast_rack
CAW95595.1
Streptococcus equi subsp equi 4047
firmicutes*
AAO82145.1
Enterococcus faecalis V583
PADR-HTH DUF1700-ahelical TM(s) Toast_rack TM
AOJ43666.1
Burkholderia thailandensis
proteobacteria
PADR-HTH DUF1700-ahelical PspC(s) LiaI-LiaF-TM Toast_rack(s)
AKD55949.1
Spirosoma radiotolerans
bacteroidetes
PADR-HTH DUF1700-ahelical MACB_PCD FTSX MACB_PCD FTSX
ACO32024.1
Acidobacterium capsulatum ATCC 511
acidobacteria, gemmatimonadetes
AOH56696.1
Bacillus muralis
PADR-HTH DUF1700-ahelical Pentapeptide
firmicutes
SIGMA-HTH PADR-HTH DUF1700-ahelical 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 Toast_rack LiaI-LiaF-TM Toast_rack SIG TM(s) HAMP HISKIN REC GerE-HTH
SIG TM(s)
firmicutes, bacteroidetes, archaea*, others
CAL82154.1
Clostridium botulinum A str ATCC 3502
LytTR LiaI-LiaF-TM Toast_rack
firmicutes
GerE-HTH REC HISKIN LiaI-LiaF-TM PspC PspC TM(s) Toast_rack
AIJ13866.1
Streptomyces lividans TK24
actinobacteria
ACU05652.1
Pedobacter heparinus DSM 2366
PspC TM(s) LiaI-LiaF-TM Toast_rack
firmicutes, bacteroidetes
SIGMA-HTH ZF TM TM LiaI-LiaF-TM TM SIG LiaI-LiaF-TM(s) Toast_rack
AEU34960.1
Granulicella mallensis MP5ACTX8
acidobacteria
ABM00816.1
Shewanella amazonensis SB2B
SIGMA-HTH TM TPR SIG Toast_rack TM Toast_rack
SIG
ANH61515.1
Dokdonia donghaensis DSW1
proteobacteria, bacteroidetes*
SIG TM(s) SIG TM(s) HISKIN REC GerE-HTH SHOCT-bihelical Toast_rack
CRY81352.1
Nocardia farcinica
actinobacteria
ALR80200.1
Streptococcus salivarius
REC GerE-HTH ABC-ATPase SIG TM(s) Toast_rack SIG TM(s) Toast_rack
firmicutes
APB74396.1
Paenibacillus polymyxa
LiaI-LiaF-TM PspA PspC coiled-coil PspA SIG TM(s) Toast_rack SIG TM HISKIN REC GerE-HTH
firmicutes
AAM36414.1
Xanthomonas citri pv citri str 306
GNTR-HTH ABC-ATPase TM(s) SIG Toast_rack SIG DUF2884
AJI31452.1
Bacillus cereus E33L
proteobacteria, firmicutes*
AGT82698.1
Amycolatopsis mediterranei RB
TraJ-RHH Toast_rack ABC-ATPase TM(s)
actinobacteria
AFY83227.1
Oscillatoria acuminata PCC 6304
SIG Toast_rack CASPASE
cyanobacteria

**Figure 6. Psp consolidated**

**A. Domain proximity network.** The network captures co-occurring domains within the top 98% 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 1**) and other commonly co-occurring domains are indicated in red and grey, respectively. **B. Phyletic spread of the most common domains.** The heatmap shows the presence 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 common partners are queried against all sequenced and completed genomes across the three major kingdoms of life. *Columns*: The major archaeal (green), bacterial (orange), and eukaryotic (blue) lineages with representative sequenced genomes. 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 S1, S2**). **C. Predominant co-occurring domain architectures in genomic neighborhoods.** UpSet plot of the most common neighboring proteins (genomic contexts >50 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
Coiled-coil
DUF3739
HISKIN
Haemagg_act
PspA
Lial-LiaF-TM
SHOCT-like
SIG
Snf7
ZNR
NLPC
PspC
Toast-rack
TM
FTSW_RODA_SPOVE
Tfu_1009
DUF1700-ahelical
PspB
PspM
MacB_PCD
DUF1707-SHOCT
DUF3046
FtsX
Pentapeptide
Query
Partner
B. Phyletic spread of most common domains
count
1500
1000
500
C. Predominant co-occurring domain architectures (in genomic neighborhoods)
Co-occurrence Counts
1661
911
607
543
514
365
291
273
256
221
208
183
163
141
138
126
122
119
117
108
107
90
81
80
78
76
74
74
73
72
72
70
69
69
65
62
62
61
LytTR
ABC-ATPase
SIG+TM+Toast-rack
TM(s)+Metallopeptidase+TM(s)
Tfu_1009
DUF1707-SHOCT+TM(s)
PspM
PspC+TM(s)
SIG+PspB
DUF1707-SHOCT+TM
Lial-LiaF-TM
SIG+TM+HAMP+HISKIN
PspC+Lial-LiaF-TM+HISKIN
PspC+TM(s)+Toast-rack
DUF1707-SHOCT+Toast-rack
clgR-HTH
Toast-rack
SIG+Lial-LiaF-TM
DO-GTPase2
TM(s)+IIGP1
PspF-NtrC-AAA+FIS-HTH
Phageshock_PspD
SIG+TM(s)+HISKIN
DUF3046
Lial-LiaF-TM+Toast-rack
TM+Toast-rack
REC+GerE-HTH
PADR-HTH
PspC+PspB
PspC
PspB
DUF1700-ahelical+TM(s)
SIG+Toast-rack
Snf7
PspA
2000
1500
1000
500
0
Prominent Domain Architectures
Co-occurring domain architectures and their frequencies

# Tables

**Table 1. Summary of Psp partner domains (query proteins)**

| Domain Name | Old Name/ Aliases | Pfam | COG | PDB |
|---|---|---|---|---|
| PspA | PspA_IM30 | PF04012 | COG1842 | 4WHE |
| Snf7 | | PF003357 | COG5491 | 5FD7 |
| PspA/Snf7 | PspA/ESCRT-III | CL0235 | | 3FRV |
| Toast_rack | DUF2154/DUF4097 | PF09922 | COG4758 | 4QRK |
| LiaI-LiaF-TM | DUF2157 | PF09925 | COG4872 | |
| PspM | Rv2743c | | | |
| PspN | Rv2742c | | | |
| PspN_N | N-terminal part, PspN | | | |
| DUF3046 | C-terminal part, PspN | PF11248 | | |
| PspB | | PF06667 | | |
| PspC | | PF04024 | COG1983 | |

# Declarations

## Acknowledgments

We are extremely 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

This work was supported by the NIH R01AI104615 grant awarded to M.L.G.; National Institutes of Health Intramural Research Program awarded to V.A. and L.A.; Oak Ridge Institute for Science and Education scholarship for the Visiting Scientist program to NIH, the Michigan State University College of Veterinary Medicine Endowed Research Funds, and 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. contributed to making figures and linking identifiers in the manuscript to reference databases. P.D. contributed to renaming some domains. J.R. and S.Z.C. built the web application.

## 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 and figures 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. doi:10.1016/j.bbamem.2004.06.010
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. doi:10.1111/j.1365-2958.2007.05893.x
3. Guest RL, Wang J, Wong JL, Raivio TL. A Bacterial Stress Response Regulates Respiratory Protein Complexes To Control Envelope Stress Adaptation. J Bacteriol. 2017;199. doi:10.1128/JB.00153-17
4. Gottesman S. Stress Reduction, Bacterial Style. J Bacteriol. 2017;199. doi:10.1128/JB.00433-17
5. Jordan S, Hutchings MI, Mascher T. Cell envelope stress response in Gram-positive bacteria. FEMS Microbiol Rev. 2008;32: 107–146. doi:10.1111/j.1574-6976.2007.00091.x
6. Raivio TL. Envelope stress responses and Gram-negative bacterial pathogenesis. Mol Microbiol. 2005;56: 1119–1128. doi:10.1111/j.1365-2958.2005.04625.x
7. Rowley G, Spector M, Kormanec J, Roberts M. Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens. Nat Rev Microbiol. 2006;4: 383–394. doi:10.1038/nrmicro1394
8. Omardien S, Drijfhout JW, van Veen H, Schachtschabel S, Riool M, Hamoen LW, et al. Synthetic antimicrobial peptides delocalize membrane bound proteins thereby inducing a cell envelope stress response. Biochim Biophys Acta Biomembr. 2018;1860: 2416–2427. doi:10.1016/j.bbamem.2018.06.005
9. Wenzel M, Kohl B, Münch D, Raatschen N, Albada HB, Hamoen L, et al. Proteomic response of Bacillus subtilis to lantibiotics reflects differences in interaction with the cytoplasmic membrane. Antimicrob Agents Chemother. 2012;56: 5749–5757. doi:10.1128/AAC.01380-12
10. Horinouchi T, Suzuki S, Hirasawa T, Ono N, Yomo T, Shimizu H, et al. Phenotypic convergence in bacterial adaptive evolution to ethanol stress. BMC Evol Biol. 2015;15: 180. doi:10.1186/s12862-015-0454-6
11. Vothknecht UC, Otters S, Hennig R, Schneider D. Vipp1: a very important protein in plastids?! J Exp Bot. 2012;63: 1699–1712. doi:10.1093/jxb/err357
12. Bultema JB, Fuhrmann E, Boekema EJ, Schneider D. Vipp1 and PspA: Related but not twins. Commun Integr Biol. 2010;3: 162–165. doi:10.4161/cib.3.2.10529
13. Westphal S, Heins L, Soll J, Vothknecht UC. 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. doi:10.1073/pnas.061501198
14. Huvet M, Toni T, Sheng X, Thorne T, Jovanovic G, Engl C, 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. doi:10.1093/molbev/msq301
15. 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. doi:10.1016/j.bbamem.2016.09.025
16. Heidrich J, Wulf V, Hennig R, Saur M, Markl J, Sönnichsen C, 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. doi:10.1074/jbc.M116.722686

17. Huvet M, Toni T, Tan H, Jovanovic G, Engl C, Buck M, et al. Model-based evolutionary analysis: the natural history of phage-shock stress response. Biochem Soc Trans. 2009;37: 762–767. doi:10.1042/BST0370762
18. 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.
19. van der Laan M, Urbanus ML, Ten Hagen-Jongman CM, Nouwen N, Oudega B, Harms N, 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. doi:10.1073/pnas.0636761100
20. Maxson ME, Darwin AJ. Identification of inducers of the Yersinia enterocolitica phage shock protein system and comparison to the regulation of the RpoE and Cpx extracytoplasmic stress responses. J Bacteriol. 2004;186: 4199–4208. doi:10.1128/JB.186.13.4199-4208.2004
21. 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. doi:10.1073/pnas.91.6.2191
22. Jovanovic G, Lloyd LJ, Stumpf MPH, Mayhew AJ, Buck M. Induction and function of the phage shock protein extracytoplasmic stress response in Escherichia coli. J Biol Chem. 2006;281: 21147–21161. doi:10.1074/jbc.M602323200
23. Mitchell AM, Silhavy TJ. Envelope stress responses: balancing damage repair and toxicity. Nat Rev Microbiol. 2019;17: 417–428. doi:10.1038/s41579-019-0199-0
24. Engl C, Beek AT, Bekker M, de Mattos JT, Jovanovic G, Buck M. Dissipation of proton motive force is not sufficient to induce the phage shock protein response in Escherichia coli. Curr Microbiol. 2011;62: 1374–1385. doi:10.1007/s00284-011-9869-5
25. McDonald C, Jovanovic G, Ces O, Buck M. Membrane Stored Curvature Elastic Stress Modulates Recruitment of Maintenance Proteins PspA and Vipp1. mBio. 2015;6: e01188-01115. doi:10.1128/mBio.01188-15
26. 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. doi:10.1101/gad.5.10.1912
27. 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. doi:10.1111/j.1365-2958.2009.06809.x
28. Brissette JL, Weiner L, Ripmaster TL, Model P. Characterization and sequence of the Escherichia coli stress-induced psp operon. J Mol Biol. 1991;220: 35–48. doi:10.1016/0022-2836(91)90379-k
29. Brissette JL, Russel M, Weiner L, Model P. Phage shock protein, a stress protein of Escherichia coli. Proc Natl Acad Sci U S A. 1990;87: 862–866. doi:10.1073/pnas.87.3.862
30. Aseeva E, Ossenbühl F, Eichacker LA, Wanner G, Soll J, Vothknecht UC. Complex formation of Vipp1 depends on its alpha-helical PspA-like domain. J Biol Chem. 2004;279: 35535–35541. doi:10.1074/jbc.M401750200
31. McDonald C, Jovanovic G, Wallace BA, Ces O, Buck M. 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. doi:10.1016/j.bbamem.2016.10.018
32. Jovanovic G, Mehta P, McDonald C, Davidson AC, Uzdavinys P, Ying L, 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. doi:10.1016/j.jmb.2013.12.016
33. Ranea JAG, Sillero A, Thornton JM, Orengo CA. Protein superfamily evolution and the last universal common ancestor (LUCA). J Mol Evol. 2006;63: 513–525. doi:10.1007/s00239-005-0289-7

34. Flores-Kim J, Darwin AJ. The Phage Shock Protein Response. Annu Rev Microbiol. 2016;70: 83–101. doi:10.1146/annurev-micro-102215-095359
35. Joly N, Engl C, Jovanovic G, Huvet M, Toni T, Sheng X, et al. Managing membrane stress: the phage shock protein (Psp) response, from molecular mechanisms to physiology. FEMS Microbiol Rev. 2010;34: 797–827. doi:10.1111/j.1574-6976.2010.00240.x
36. Domínguez-Escobar J, Wolf D, Fritz G, Höfler C, Wedlich-Söldner R, Mascher T. Subcellular localization, interactions and dynamics of the phage-shock protein-like Lia response in Bacillus subtilis. Mol Microbiol. 2014;92: 716–732. doi:10.1111/mmi.12586
37. Datta P, Ravi J, Guerrini V, Chauhan R, Neiditch MB, Shell SS, et al. The Psp system of Mycobacterium tuberculosis integrates envelope stress-sensing and envelope-preserving functions. Mol Microbiol. 2015;97: 408–422. doi:10.1111/mmi.13037
38. Ravi J, Anantharaman V, Aravind L, Gennaro ML. Variations on a theme: evolution of the phage-shock-protein system in Actinobacteria. Antonie Van Leeuwenhoek. 2018;111: 753–760. doi:10.1007/s10482-018-1053-5
39. Mascher T, Zimmer SL, Smith T-A, Helmann JD. Antibiotic-inducible promoter regulated by the cell envelope stress-sensing two-component system LiaRS of Bacillus subtilis. Antimicrob Agents Chemother. 2004;48: 2888–2896. doi:10.1128/AAC.48.8.2888-2896.2004
40. Otters S, Braun P, Hubner J, Wanner G, Vothknecht UC, Chigri F. The first α-helical domain of the vesicle-inducing protein in plastids 1 promotes oligomerization and lipid binding. Planta. 2013;237: 529–540. doi:10.1007/s00425-012-1772-1
41. Helmann JD. Bacillus subtilis extracytoplasmic function (ECF) sigma factors and defense of the cell envelope. Curr Opin Microbiol. 2016;30: 122–132. doi:10.1016/j.mib.2016.02.002
42. Eiamphungporn W, Helmann JD. The Bacillus subtilis sigma(M) regulon and its contribution to cell envelope stress responses. Mol Microbiol. 2008;67: 830–848. doi:10.1111/j.1365-2958.2007.06090.x
43. Jordan S, Junker A, Helmann JD, Mascher T. 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. doi:10.1128/JB.00310-06
44. Darwin AJ. The phage-shock-protein response. Mol Microbiol. 2005;57: 621–628. doi:10.1111/j.1365-2958.2005.04694.x
45. 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. doi:10.1128/jb.182.2.311-319.2000
46. Elderkin S, Bordes P, Jones S, Rappas M, Buck M. Molecular determinants for PspA-mediated repression of the AAA transcriptional activator PspF. J Bacteriol. 2005;187: 3238–3248. doi:10.1128/JB.187.9.3238-3248.2005
47. Yamaguchi S, Reid DA, Rothenberg E, Darwin AJ. 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. doi:10.1111/mmi.12122
48. Lloyd LJ, Jones SE, Jovanovic G, Gyaneshwar P, Rolfe MD, Thompson A, 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. doi:10.1074/jbc.M408994200
49. Green RC, Darwin AJ. PspG, a new member of the Yersinia enterocolitica phage shock protein regulon. J Bacteriol. 2004;186: 4910–4920. doi:10.1128/JB.186.15.4910-4920.2004
50. Wolf D, Kalamorz F, Wecke T, Juszczak A, Mäder U, Homuth G, et al. In-depth profiling of the LiaR response of Bacillus subtilis. J Bacteriol. 2010;192: 4680–4693. doi:10.1128/JB.00543-10
51. Suntharalingam P, Senadheera MD, Mair RW, Lévesque CM, Cvitkovitch DG. The LiaFSR

system regulates the cell envelope stress response in Streptococcus mutans. J Bacteriol. 2009;191: 2973–2984. doi:10.1128/JB.01563-08
52. Kesel S, Mader A, Höfler C, Mascher T, Leisner M. Immediate and heterogeneous response of the LiaFSR two-component system of Bacillus subtilis to the peptide antibiotic bacitracin. PloS One. 2013;8: e53457. doi:10.1371/journal.pone.0053457
53. Schrecke K, Jordan S, Mascher T. Stoichiometry and perturbation studies of the LiaFSR system of Bacillus subtilis. Mol Microbiol. 2013;87: 769–788. doi:10.1111/mmi.12130
54. Kleine B, Chattopadhyay A, Polen T, Pinto D, Mascher T, Bott M, et al. The three-component system EsrISR regulates a cell envelope stress response in Corynebacterium glutamicum. Mol Microbiol. 2017;106: 719–741. doi:10.1111/mmi.13839
55. Weiner L, Brissette JL, Ramani N, Model P. Analysis of the proteins and cis-acting elements regulating the stress-induced phage shock protein operon. Nucleic Acids Res. 1995;23: 2030–2036. doi:10.1093/nar/23.11.2030
56. Srivastava D, Moumene A, Flores-Kim J, Darwin AJ. Psp Stress Response Proteins Form a Complex with Mislocalized Secretins in the Yersinia enterocolitica Cytoplasmic Membrane. mBio. 2017;8. doi:10.1128/mBio.01088-17
57. DeAngelis CM, Nag D, Withey JH, Matson JS. Characterization of the Vibrio cholerae Phage Shock Protein Response. J Bacteriol. 2019;201. doi:10.1128/JB.00761-18
58. Anh Le TT, Thuptimdang P, McEvoy J, Khan E. Phage shock protein and gene responses of Escherichia coli exposed to carbon nanotubes. Chemosphere. 2019;224: 461–469. doi:10.1016/j.chemosphere.2019.02.159
59. 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. doi:10.1111/mmi.12380
60. Jordan S, Rietkötter E, Strauch MA, Kalamorz F, Butcher BG, Helmann JD, et al. LiaRS-dependent gene expression is embedded in transition state regulation in Bacillus subtilis. Microbiol Read Engl. 2007;153: 2530–2540. doi:10.1099/mic.0.2007/006817-0
61. Manganelli R, Gennaro ML. Protecting from Envelope Stress: Variations on the Phage-Shock-Protein Theme. Trends Microbiol. 2017;25: 205–216. doi:10.1016/j.tim.2016.10.001
62. Zhang L, Kato Y, Otters S, Vothknecht UC, Sakamoto W. Essential role of VIPP1 in chloroplast envelope maintenance in Arabidopsis. Plant Cell. 2012;24: 3695–3707. doi:10.1105/tpc.112.103606
63. Armstrong RM, Adams KL, Zilisch JE, Bretl DJ, Sato H, Anderson DM, et al. Rv2744c Is a PspA Ortholog That Regulates Lipid Droplet Homeostasis and Nonreplicating Persistence in Mycobacterium tuberculosis. J Bacteriol. 2016;198: 1645–1661. doi:10.1128/JB.01001-15
64. Castelle CJ, Banfield JF. Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. Cell. 2018;172: 1181–1197. doi:10.1016/j.cell.2018.02.016
65. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. Nat Microbiol. 2016;1: 16048. doi:10.1038/nmicrobiol.2016.48
66. 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. doi:10.1046/j.1365-2958.2001.02235.x
67. 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. doi:10.1111/j.1365-2958.2006.05047.x
68. Becker LA, Bang I-S, Crouch M-L, Fang FC. 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. doi:10.1111/j.1365-2958.2005.04604.x
69. Caillat C, Maity S, Miguet N, Roos WH, Weissenhorn W. The role of VPS4 in ESCRT-III polymer remodeling. Biochem Soc Trans. 2019;47: 441–448. doi:10.1042/BST20180026
70. Schöneberg J, Pavlin MR, Yan S, Righini M, Lee I-H, Carlson L-A, et al. ATP-dependent force generation and membrane scission by ESCRT-III and Vps4. Science. 2018;362: 1423–1428. doi:10.1126/science.aat1839
71. 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. doi:10.1042/BJ20090957
72. Godinho RM da C, Crestani J, Kmetzsch L, Araujo G de S, Frases S, Staats CC, 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. doi:10.1038/srep06198
73. Webster BM, Colombi P, Jäger J, Lusk CP. Surveillance of nuclear pore complex assembly by ESCRT-III/Vps4. Cell. 2014;159: 388–401. doi:10.1016/j.cell.2014.09.012
74. Lee I-H, Kai H, Carlson L-A, Groves JT, Hurley JH. 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. doi:10.1073/pnas.1518765113
75. 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. doi:10.1128/JB.01250-12
76. 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. doi:10.1111/j.1365-2958.2009.06885.x
77. Wallrodt I, Jelsbak L, Thomsen LE, Brix L, Lemire S, Gautier L, 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. doi:10.1099/jmm.0.072223-0
78. Burki F, Shalchian-Tabrizi K, Pawlowski J. Phylogenomics reveals a new “megagroup” including most photosynthetic eukaryotes. Biol Lett. 2008;4: 366–369. doi:10.1098/rsbl.2008.0224
79. Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, 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. doi:10.1073/pnas.0807880106
80. Osadnik H, Schöpfel M, Heidrich E, Mehner D, Lilie H, Parthier C, 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. doi:10.1111/mmi.13154
81. Tang S, Henne WM, Borbat PP, Buchkovich NJ, Freed JH, Mao Y, et al. Structural basis for activation, assembly and membrane binding of ESCRT-III Snf7 filaments. eLife. 2015;4. doi:10.7554/eLife.12548
82. Zhang L, Kondo H, Kamikubo H, Kataoka M, Sakamoto W. VIPP1 Has a Disordered C-Terminal Tail Necessary for Protecting Photosynthetic Membranes against Stress. Plant Physiol. 2016;171: 1983–1995. doi:10.1104/pp.16.00532
83. Anantharaman V, Aravind L. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome Biol. 2003;4: R11. doi:10.1186/gb-2003-4-2-r11
84. Wollert T, Wunder C, Lippincott-Schwartz J, Hurley JH. Membrane scission by the ESCRT-III

complex. Nature. 2009;458: 172–177. doi:10.1038/nature07836
85. Fiuza M, Letek M, Leiba J, Villadangos AF, Vaquera J, Zanella-Cléon I, et al. Phosphorylation of a novel cytoskeletal protein (RsmP) regulates rod-shaped morphology in Corynebacterium glutamicum. J Biol Chem. 2010;285: 29387–29397. doi:10.1074/jbc.M110.154427
86. Lee YH, Kingston AW, Helmann JD. Glutamate dehydrogenase affects resistance to cell wall antibiotics in Bacillus subtilis. J Bacteriol. 2012;194: 993–1001. doi:10.1128/JB.06547-11
87. Jovanovic G, Mehta P, Ying L, Buck M. Anionic lipids and the cytoskeletal proteins MreB and RodZ define the spatio-temporal distribution and function of membrane stress controller PspA in Escherichia coli. Microbiol Read Engl. 2014;160: 2374–2386. doi:10.1099/mic.0.078527-0
88. Ballas SK, Mohandas N, Marton LJ, Shohet SB. Stabilization of erythrocyte membranes by polyamines. Proc Natl Acad Sci U S A. 1983;80: 1942–1946. doi:10.1073/pnas.80.7.1942
89. Lee M-J, Huang C-Y, Sun Y-J, Huang H. Cloning and characterization of spermidine synthase and its implication in polyamine biosynthesis in Helicobacter pylori strain 26695. Protein Expr Purif. 2005;43: 140–148. doi:10.1016/j.pep.2005.04.017
90. Cerutti ML, Otero LH, Smal C, Pellizza L, Goldbaum FA, Klinke S, et al. Structural and functional characterization of a cold adapted TPM-domain with ATPase/ADPase activity. J Struct Biol. 2017;197: 201–209. doi:10.1016/j.jsb.2016.10.010
91. Eberhardt RY, Bartholdson SJ, Punta M, Bateman A. The SHOCT domain: a widespread domain under-represented in model organisms. PloS One. 2013;8: e57848. doi:10.1371/journal.pone.0057848
92. Miller C, Kong J, Tran TT, Arias CA, Saxer G, Shamoo Y. Adaptation of Enterococcus faecalis to daptomycin reveals an ordered progression to resistance. Antimicrob Agents Chemother. 2013;57: 5373–5383. doi:10.1128/AAC.01473-13
93. Boyle-Vavra S, Yin S, Jo DS, Montgomery CP, Daum RS. VraT/YvqF is required for methicillin resistance and activation of the VraSR regulon in Staphylococcus aureus. Antimicrob Agents Chemother. 2013;57: 83–95. doi:10.1128/AAC.01651-12
94. Gardete S, Wu SW, Gill S, Tomasz A. Role of VraSR in antibiotic resistance and antibiotic-induced stress response in Staphylococcus aureus. Antimicrob Agents Chemother. 2006;50: 3424–3434. doi:10.1128/AAC.00356-06
95. Wolf D, Domínguez-Cuevas P, Daniel RA, Mascher T. Cell envelope stress response in cell wall-deficient L-forms of Bacillus subtilis. Antimicrob Agents Chemother. 2012;56: 5907–5915. doi:10.1128/AAC.00770-12
96. Hauf S, Möller L, Fuchs S, Halbedel S. PadR-type repressors controlling production of a non-canonical FtsW/RodA homologue and other trans-membrane proteins. Sci Rep. 2019;9: 10023. doi:10.1038/s41598-019-46347-w
97. Zhang C, Nietfeldt J, Zhang M, Benson AK. Functional consequences of genome evolution in Listeria monocytogenes: the lmo0423 and lmo0422 genes encode sigmaC and LstR, a lineage II-specific heat shock system. J Bacteriol. 2005;187: 7243–7253. doi:10.1128/JB.187.21.7243-7253.2005
98. Wang L, Xiao S, Chen X, Chen S, Wang S, Wang C, et al. ytiB and ythA Genes Reduce the Uranium Removal Capacity of Bacillus atrophaeus. Int J Mol Sci. 2019;20. doi:10.3390/ijms20071766
99. Zhang L, Sakamoto W. Possible function of VIPP1 in maintaining chloroplast membranes. Biochim Biophys Acta. 2015;1847: 831–837. doi:10.1016/j.bbabio.2015.02.013
100. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res. 2013;41: D36-42. doi:10.1093/nar/gks1195
101. Tatusova T, Ciufo S, Federhen S, Fedorov B, McVeigh R, O'Neill K, et al. Update on RefSeq microbial genomes resources. Nucleic Acids Res. 2015;43: D599-605.

doi:10.1093/nar/gku1062
102. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. 2014;42: D581-591. doi:10.1093/nar/gkt1099
103. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res. 2017;45: D535–D542. doi:10.1093/nar/gkw1017
104. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44: D733-745. doi:10.1093/nar/gkv1189
105. Federhen S. The NCBI Taxonomy database. Nucleic Acids Res. 2012;40: D136-143. doi:10.1093/nar/gkr1178
106. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25: 3389–3402. doi:10.1093/nar/25.17.3389
107. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421. doi:10.1186/1471-2105-10-421
108. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005;33: W244-248. doi:10.1093/nar/gki408
109. Nielsen H. Predicting Secretory Proteins with SignalP. Methods Mol Biol Clifton NJ. 2017;1611: 59–73. doi:10.1007/978-1-4939-7015-5_6
110. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305: 567–580. doi:10.1006/jmbi.2000.4315
111. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004;338: 1027–1036. doi:10.1016/j.jmb.2004.03.016
112. 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. doi:10.1093/nar/gkm256
113. Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. Nucleic Acids Res. 2008;36: W197-201. doi:10.1093/nar/gkn238
114. Mistry J, Finn R. Pfam: a domain-centric method for analyzing proteins and proteomes. Methods Mol Biol Clifton NJ. 2007;396: 43–58. doi:10.1007/978-1-59745-515-2_4
115. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39: W29-37. doi:10.1093/nar/gkr367
116. Lassmann T. Kalign 3: multiple sequence alignment of large data sets. Bioinforma Oxf Engl. 2019. doi:10.1093/bioinformatics/btz795
117. Lassmann T, Sonnhammer ELL. Kalign--an accurate and fast multiple sequence alignment algorithm. BMC Bioinformatics. 2005;6: 298. doi:10.1186/1471-2105-6-298
118. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32: 1792–1797. doi:10.1093/nar/gkh340
119. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5: 113. doi:10.1186/1471-2105-5-113
120. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PloS One. 2010;5: e9490. doi:10.1371/journal.pone.0009490
121. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. shiny: Web Application Framework for R. 2019. Available: https://CRAN.R-project.org/package=shiny

122. 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. Available: http://shiny.rstudio.com/
123. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available: https://www.R-project.org/
124. RStudio Team. RStudio: Integrated Development Environment for R. Boston, MA: RStudio, PBC.; 2020. Available: http://www.rstudio.com/
125. Mike Bostock CY Kerry Rodden, Kevin Warne, Kent Russell, Florian Breitwieser. Sunburst "Htmlwidget." 2020. Available: https://github.com/timelyportfolio/sunburstR
126. 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. Available: http://haozhu233.github.io/kableExtra/, https://github.com/haozhu233/kableExtra
127. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: Visualization of Intersecting Sets. IEEE Trans Vis Comput Graph. 2014;20: 1983–1992. doi:10.1109/TVCG.2014.2346248
128. Enrico Bonatesta UB Christoph Horejs-Kainrath. Multiple Sequence Alignment. 2020. Available: http://www.bioinf.jku.at/software/msa/
129. 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. Available: http://ape-package.ird.fr/
130. Fellows I. Word Clouds. 2018. Available: http://blog.fellstat.com/?cat=11
131. Dawei Lang GC. Create Word Cloud by 'htmlwidget. 2018. Available: https://github.com/lchiffon/wordcloud2
132. Jake Conway NG. A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets. 2019. Available: http://github.com/hms-dbmi/UpSetR
133. Müller K. here: A Simpler Way to Find Your Files. 2017. Available: https://CRAN.R-project.org/package=here
134. 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. J Open Source Softw. 2019;4: 1686. doi:10.21105/joss.01686
135. Martin Morgan MR. Access the Bioconductor Project Package Repository. 2019. Available: https://github.com/Bioconductor/BiocManager
136. Perry M. Flexible Heatmaps for Functional Genomics and Sequence Features. 2016. Available: https://git.bioconductor.org/packages/heatmaps
137. Yihui Xie JJ Allaire, Grolemund G. R Markdown: The Definitive Guide. Boca Raton, Florida: Chapman and Hall/CRC; 2018. Available: https://bookdown.org/yihui/rmarkdown