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

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

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

*Corresponding authors:

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

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

Abstract

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

Importance

Phage shock proteins (PSP) are virulence-associated, cell membrane stress-protective systems. They have mostly been characterized in proteobacteria and firmicutes. We now show that PSP systems present in the Last Universal Common Ancestor, and that homologs have evolved and diversified into newly identified functional contexts. Recognizing the conservation and evolution of PSP systems across bacterial phyla contributes new knowledge to our current understanding of

stress response mechanisms in prokaryotes. Moreover, the newly discovered PSP modularity will likely prompt new studies of lineage-specific cell-envelope structures, lifestyles, and adaptation mechanisms. Finally, our results validate use of domain architecture and genetic context for discovery in comparative genomics.

Keywords

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

Introduction

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

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

PSP expression is regulated by the peripheral membrane protein PspA and the homologous plant protein Vipp1 (11–16, 23–31), which both contain the coiled-coil structure PspA domain (16, 17, 19, 20, 32, 33) (**Table S1**). Bacterial PspA senses and responds to extracellular stress (14, 23, 33–35).

71 Its domain, Pfam PspA_IM30 (inner membrane 30KDa domain) is found in bacteria, photosynthetic
72 eukaryotes, and archaea (15–17, 33, 36, 37) (**Fig. 1A**). Proteins containing PspA_IM30 are known to
73 facilitate membrane fusion (16) and prevent leakage of protons through the membrane (2, 23) in
74 response to a wide range of surface stresses, including changes in membrane potential, *ΔpH* (2,
75 23) and PMF (38–41), (l)antibiotic stress (7), heat/osmotic shock (42) and mechanical stress via
76 stored curvature elastic stress (13, 43–45). Vipp1 is involved in thylakoid biogenesis, vesicular lipid
77 transport and storage, and membrane shaping (16, 26, 27, 32).
78
79 The classical PSP system first discovered in Proteobacteria (46) is not conserved across all bacterial
80 and archaeal clades (12, 33, 47). However, the PSP systems described in Firmicutes and
81 Actinobacteria (3, 11, 14, 15, 35, 36, 48, 49) include clade-specific PSP proteins with distinct
82 domain architectures and conserved functions (*e.g.,* transcriptional regulation, membrane tethering
83 and relocation). Several key questions remain, including how PspA evolved to function with distinct
84 partners across diverse lineages, about the identities of other integral components of the PSP
85 system beyond the classical systems and complete characterization of cognate partners, if any
86 homologs participate in novel functions, and whether common themes emerge from the phyletic
87 patterns of PSP systems.
88
89 We set out to evaluate the evolutionary and functional significance of the PSP system through a
90 comprehensive analysis of PspA occurrences, associated genomic contexts, and phyletic spreads
91 across the three superkingdoms or domains of life (50, 51). We resolved PSP proteins into
92 constituent domain architectures and queried each individual domain in our comparative genomics
93 and phylogenetic analyses. This domain-centric approach is ideal to address the questions posed
94 above. We identified several novel players in the PSP envelope stress response system and
95 described their putative function and evolution. The PSP repertoire — including new members,

homologs, domain architectures, genomic neighborhoods, phyletic spreads, and phylogenetic analyses in all representative species — is available for researchers to systematically explore through tabular and graphical summaries on our user-friendly, interactive companion web app.

Results and Discussion

The PSP system presents an evolutionary and functional paradox: PspA appears across the tree of life but exists in distinct genomic neighborhoods and with different functional partners (15, 30, 52). We, therefore, sought to identify and characterize all homologs of PspA and cognate partner domains across a dataset representative of the tree of life. We used comparative genomics and a computational molecular evolutionary approach to comprehensively identify PSP homologs. We discovered several novel conserved, lineage-specific domain architectures and genomic contexts of PSP systems and delineated their phyletic spreads across the tree of life in the process (see *Methods* for details). All our findings (data, results, and visual summaries) are available in an interactive and searchable web application (jravilab.shinyapps.io/psp_evolution) referred to henceforth as the 'web app.' The web app can be queried with the protein accession numbers cited inline throughout the *Results* section.

Phyletic spread of classical Psp components across the tree of life

Describing the PSP network of proteins across the tree of life requires characterization of classical PSP components in Proteobacteria, Actinobacteria, and Firmicutes. Among Proteobacteria, we analyzed the *E. coli psp* operon, which is widely conserved across Gammaproteobacteria (10, 12, 39, 46, 53–55). PspA, PspB, and PspC are each composed of their namesake domains spanning almost their entire protein lengths. The transcriptional regulator PspF is encoded on the opposite strand and shares a central promoter region with the *pspABC* operon in most

120 Gammaproteobacteria, and only with *pspA* in other organisms. PspF contains an enhancer-binding,
121 specialized AAA-ATPase NtrC-AAA domain, in addition to the Fis-type HTH (helix-turn-helix) DNA-
122 binding domain. In Actinobacteria, the *M. tuberculosis psp* operon (14, 15, 36) includes the
123 namesake PspA protein; a transcription factor, ClgR, containing a cro-like-HTH (cHTH) DNA-
124 binding domain unique to actinobacteria; an integral membrane protein, Rv2743c, distinct from *E.*
125 *coli* PspC and PspB; and Rv2742c, an uncharacterized protein carrying a "Domain of Unknown
126 Function,'' DUF3046 (14, 15, 36). For consistency, we renamed Rv2743c and Rv2742c PspM and
127 PspN, respectively (Datta *et al.*, unpublished). *B. subtilis,* a member of Firmicutes, contains the *lia*
128 operon. LiaI is a small transmembrane (TM) protein (11, 49, 56), LiaF contains N-terminal TM helices
129 (DUF2157) and a C-terminal globular domain (DUF2154), and LiaG contains an uncharacterized
130 domain DUF4097. When we analyzed the domain architectures of these operons, we defined two
131 novel domains of the PSP system – the LiaI-LiaF-TM sensory domain that unifies four TM domains
132 (4TM) and DUF2157 from LiaI/LiaF, as well as Toastrack, unifying DUF4097/DUF2154/DUF2807
133 Pfam models. The *lia* operon is controlled by a two-component system LiaRS (11, 35). LiaR contains
134 a receiver domain, and a winged HTH (wHTH) DNA-binding domain (REC-wHTH), while LiaS is a
135 sensor kinase with a TM region, intracellular HAMP, and a histidine kinase (HISKIN) signaling
136 domain (57). Information on domains in this section is presented in **Text S1.1** and **Table S1**, and
137 graphically in **Fig. 1A** and **1B**.
138

139 ## PspA

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

144 (**Fig. 1C**). Analysis using multi-query sequence searches, we also identified Snf7 proteins from the
145 eukaryotic membrane-remodeling ESCRT systems **(18, 63–67)** as homologs (**Fig. 1B**). Snf7 was
146 then used for iterative searches, which yielded several pan-eukaryotic and archaeal homologs
147 distinct from Vipp1 (**Fig. 1B** and **1C**; **Table S3**), supporting a considerably wider evolutionary
148 distribution for PspA, corroborated by other recent reports (19, 20).
149

150 Toastrack and LiaI-LiaF-TM

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

160

161 PspB and PspC

162 **PspB**, often part of the proteobacterial *psp* operon, is another integral membrane protein implicated
163 in stress sensing and virulence (58–60). PspB has an N-terminal anchoring TM helix followed by an
164 α-helical cytoplasmic globular domain. While PspB is rarely found outside proteobacteria (**Fig. 1C**),
165 we discovered previously unrecognized divergent PspB domains fused to PspC (see section on
166 *PspC architectures)*.
167

168 **PspC** was first identified in the proteobacterial PSP system; it is critical in sensing membrane stress
169 and restoring envelope integrity. Recent studies demonstrate that PspC in some bacterial species
170 may function independently of PspA in response to certain membrane stressors (59, 61–63).
171 Another integral membrane protein, our analysis showed PspC has two TM helices, the first being
172 a cryptic TM ('MSA' tab, web app). A conserved arginine (a positively charged residue) found
173 between the two TM regions is reminiscent of structures serving in membrane-associated stimulus
174 sensing, which may provide clues towards the biological roles of PspC. We observed that the PspC
175 domain is pan-bacterial and also present in a few archaeal clades (**Fig. 1C**). Our analysis showed
176 that PspC has two TM helices, the first being a cryptic TM. We also identified several novel domain
177 architectures and genomic contexts of PspC (see section on *PspC architectures*).
178

179 ## PspM and PspN

180 Domain-centric searches found no discernible homologs for PspMN outside Actinobacteria (**Fig.**
181 **1C**; web app), as with prior full protein searches (15, 36, 64). Two TM helices make up the
182 corynebacterial integral membrane partner **PspM** (encoded by *rv2743c*). Our analyses revealed
183 minimal PspM variation by MSA ('MSA' tab, web app) and a narrow phyletic spread restricted to
184 mycobacteria and corynebacteria (**Fig. 1B**) (15, 36, 64).
185

186 The fourth member in the Mycobacterial operon, PspN (Rv2742c), contains a C-terminal DUF3046
187 and an N-terminal domain we call **PspN_N** (**Table S1**; **Text S1.2**; 'MSA' tab in the web app) (15,
188 36, 64). **DUF3046** is prevalent only in actinobacteria (**Fig. 1C**). Moreover, the *M. tuberculosis*
189 genome carries a second copy of DUF3046 (*rv2738c*), located four genes downstream of *pspN*. We
190 infer that Rv2738c, rather than *pspN*, contains the ancestral copy of DUF3046 (**Text S1.2**). DUF3046
191 found as part of PspN is likely a duplicate translocated into the PspN open reading frame (ORF) of

192 mycobacteria, especially the *M. tuberculosis* complex. Further, unlike the pan-actinobacterial
193 DUF3046, the N-terminal region, PspN_N, is conserved only in *M. tuberculosis*, with remnants of
194 the coding region existing as potential pseudo-genes or degraded sequences in a few closely
195 related mycobacteria (*e.g., M. avium*).
196

197 Evolution of PspA

198 After identifying eukaryotic Snf7 as homologous, we sought to expand on the history of PspA. We
199 used a structure-informed MSA and phylogenetic tree ('MSA' tab, web app; **Fig. 2; Table S1**) to
200 investigate the PspA/Snf7 superfamily, and define its evolutionary origins.
201

202 Identifying clades carrying PspA and Snf7

203 PspA+ clades comprise all versions of the PspA domain phylogenetically similar to the 'classical'
204 PspA domain rather than the Snf7 domains. We found that most bacterial lineages contain PspA+
205 clade members (**Fig. 1C**), with a few transfers to archaea. Among eukaryotes, only those containing
206 plastids and two flagella (comprising the SAR/HA supergroup and excavata) have members of this
207 clade (**Fig. 1C**; **Fig. 2**; web app) (65, 66).
208

209 Similarly, we found that the curated set of PspA-like proteins contains a divergent cluster of proteins
210 belonging to what we now call the Snf7+ clade (**Fig. 2**; web app). Snf7 is part of the ESCRT-III
211 complex required for endosome-mediated trafficking via multivesicular body formation and sorting
212 and has a predominantly archaeo-eukaryotic phyletic pattern (**Fig. 1C**) (18, 67–71).
213

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

We explored the evolution of the PspA-Snf7 superfamily using an MSA of comprehensively selected PspA/Snf7 homologs from distinct clades across the tree of life (50, 51) and different domain architectures (**Fig. 2**; web app). The MSA revealed a unique insertion of heptad repeats in actinobacterial PspA sequences, likely a lineage-specific adaptation. A few cyanobacterial PspA homologs similar to the ancestral plant variant Vipp1 also contained a C-terminal extension (16, 32, 72). A maximum-likelihood phylogenetic tree was generated using this MSA (**Fig. 2**; web app), which showed clear clustering into i) easily distinguishable clades of homologs from Actinobacteria, Firmicutes, and Proteobacteria; ii) the eukaryotic Vipp1 clade grouped with cyanobacterial homologs indicating plastid origins; and iii) Snf7 domain-containing homologs from archaea and eukaryotes forming a well-defined, separate branch. Together, these observations support that the PspA/Snf7 superfamily is ubiquitous across the superkingdoms/domains of life and originated from a PspA-like gene present in the last universal common ancestor (LUCA), as independently corroborated by recent studies (19, 20).

PspA: Novel architectures and neighborhoods

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

Domain Architectures

Our searches revealed that the underlying domain architectures of most PspA homologs (>98%; ~2500 homologs) show little variation (web app). In most lineages, PspA homologs contain only the

238 characteristic PspA_IM30 domain (**Fig. 3**; **Table S3;** web app). The remaining <2% of homologs
239 contain PspA_IM30 repeats in tandem or fusions with other domains, including NlpC/P60 and a
240 novel PspA-associated (PspAA) domain (**Fig. 3**; **Text S2.1**). The PspA_IM30–NlpC/P60 hydrolase
241 fusion is predicted to catalyze the modification of phosphatidylcholine, likely altering membrane
242 composition (**Fig. 3**; **Table S3**; AFZ52345.1) (73). Similarly, Snf7 showed minimal variation in domain
243 architecture, with occasional fusions (<5%) in eukaryotes, with some potentially attributable to
244 genome annotation errors. Some actinobacteria have an Snf7 homolog fused to an RND-family
245 lipid/fatty acid transporter, flanked by two genes encoding a *Mycobacterium*-specific TM protein
246 with a C-terminal cysteine-rich domain (74).
247
248 *Paralogs*
249 We found that genomes containing multiple copies of PspA do not maintain domain architecture
250 and genomic context among paralogs (**Fig. 3**; **Table S3**; 'Phylogeny –> Paralog' tab, web app). In
251 cyanobacteria, dyads/triads of PspA paralogs occur as adjacent, occasionally fused genes (**Fig. 3**;
252 **Table S3**; *e.g.,* BAG06015.1). Consistent with earlier reports (25, 27), we observed that these
253 neighboring PspA genes are part of two distinct clusters of homologs, one resembling the bacterial
254 PspA and the other the eukaryotic (plastid) Vipp1 (**Fig. 3**). Snf7 gene clusters also appear
255 occasionally in archaeal and eukaryotic species (**Table S3**; *e.g.* CBY21170.1). This suggests
256 tandem gene duplication and may relate to the role of these clusters in membrane stabilization as
257 oligomeric complexes, as in the eukaryotic ESCRT systems (75). Further expansion on PspA/Snf7
258 paralogs and their likely evolution, including inference of gene duplication *vs.* horizontal gene
259 transfer from domain architectures and genomic contexts, is available in the web app. Occasionally,
260 PspA and Snf7 co-occur in conserved gene neighborhoods (**Fig. 3**; **Table S3**) such as in
261 deltaproteobacteria (AKJ06548.1); in bacteroidetes (CBH24266.1), these neighborhoods also

262 contain a third gene encoding a coiled-coil protein with a structure reminiscent of the PspA coiled-
263 coils (**Table S3**).
264

265 Novel variations of classical Psp genomic contexts

266 We next explored the gene-neighborhood themes for each Psp member (**Fig. 3**; **Table S3**). Major
267 themes representing variations of classical Psp operons are summarized below. For the full list of
268 genomic contexts and phyletic spreads, see web app: 'Data,' 'Genomic Contexts' tabs.
269

270 *Diversity of the* pspFABC *operon*

271 In addition to proteobacteria, the classic *pspABC* occurs in nitrospirae and some spirochaetes (**Fig.**
272 **3**; **Table S3**). There are also variations to this theme (web app): 1) PspC is fused to a divergent C-
273 terminal PspB in addition to a single PspB in the operon (**Fig. 3**; **Table S3**; *e.g.,* ANW39986.1); 2)
274 PspB duplicates appear in the operon (**Fig. 3**; **Table S3**; *e.g.,* AOL22920.1); 3) PspD occurs along
275 with this operon only in gammaproteobacteria (**Fig. 3**; **Table S3**; *e.g.,* ANW39986.1). Key variations
276 in the transcription regulator PspF (NtrC-AAA and HTH) include: 1) an operon with additional genes
277 for PspB and PspC fusions and Toastrack-containing proteins in a few spirochaetes; 2) operons in
278 gammaproteobacteria with genes encoding NtrC-like transcription factors with N-terminal ligand-
279 binding ACT and PAS domains (**Fig. 3**; **Table S3**; *e.g.,* AEN65073.1) (76), and a further protein of
280 the DO-GTPase family, predicted to play a role in membrane-related stresses (77, 78), which
281 occasionally feature an additional PspF; and 3) in some alphaproteobacteria, the PspC in PspFABC
282 has been replaced by multiple Toastrack-containing proteins (**Fig. 3**; **Table S3**; *e.g.,* ANQ40502.1).
283 Thus, classic *pspABC* presents variation in structure and encoded function.
284

285 *Associations with Vps4 and other classical AAA+-ATPases*

286 The core of an ESCRT complex in archaea is defined by the co-occurrence of one or more copies
287 of Snf7 (*e.g.,* OLS27540.1; [**Table S3**]) and a gene encoding a **VPS4-like AAA+-ATPase** (with an
288 N-terminal MIT and C-terminal oligomerization domain [**Table S2**] (79). In our analysis, we observed
289 diversity in *vps4* gene neighborhoods (**Text S2.2**). ESCRT complexes are well-studied in
290 eukaryotes, but less so in archaea, where they were originally associated with cell division and
291 named for this role (Cdv system) (80). Crenarchaeota and Thaumarchaeota contain *cdvABC*. *cdvA*
292 has no eukaryotic homolog, while *cdvB* and *cdvC* encode homologs of ESCRT-III proteins and
293 Vps4, respectively. Recent studies have shown that eukaryotes originate within the archaeal clade
294 of Asgardarchaeota (80). Consistent with this, Asgardarchaeota show a conserved gene
295 neighborhood coupling Snf7, a VPS4-like AAA+ ATPase gene, and a gene coding for ESCRT-II
296 wHTH-domain (81) corresponding to the cognate eukaryotic ESCRT complex. These operons may
297 be further extended with additional copies of Snf7 genes and other genes coding for a TM protein
298 and an ABC ATPase. Additionally, examples of transfers of VPS4-like AAA+-ATPase from archaea
299 to several bacterial lineages were observed (**Table S3**; *e.g.,* ACB74714.1). In these cases, *snf7* is
300 displaced by an unrelated gene containing TPR repeat and 6TM, suggesting membrane proximity.
301 The bacterial PspA (**Table S3**; *e.g.* AEY64321.1) may occur with a distinct protein containing two
302 AAA+-ATPase domains – with the N-terminal copy inactive – in some bacterial clades (**Text S2.2**).
303 Both PspA and the membrane-associated Snf7, along with AAA+-ATPase, may occur in longer
304 operons with other genes encoding an ABC-ATPase, permease, and a solute- and peptidoglycan-
305 binding protein with PBPB-OmpA domains (**Table S3**; *e.g.,* OGG56892.1). These associations
306 between certain bacterial PspA and Snf7 clade members with AAA+-ATPases suggests a role in
307 ATP-dependent membrane remodeling (**Fig. 3**; **Table S3**) comparable to Snf7 and VPS4 in the
308 ESCRT-III-mediated membrane remodeling in the archaeo-eukaryotic lineage (19, 20, 67–69, 71,

309 75). The additional transporter components seen in some of these systems suggest a link between
310 this membrane remodeling and solute transport. Here the OmpA domain could interact with
311 peptidoglycan, while the PBPB domain binds the solute trafficked by the linked ABC transporter.
312 Finally, recent structural studies have reached similar evolutionary conclusions, i.e., that Vipp1,
313 PspA, and Snf7 proteins arose from a common ancestor and can play similar roles (19, 20).
314

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

316 We discovered two novel overlapping genomic associations across various bacteria linking PspA
317 with the Band-7 domain and CesT/Tir-like chaperone domain proteins (**Fig. 3**; **Tables S2** and **S3**;
318 'MSA' tab, web app). Band-7 has been implicated in macromolecular complex assembly, playing a
319 chaperone-like role in binding peptides (82), and in both membrane-localization and membrane
320 dynamics (83). CesT/Tir chaperone domains have been shown to mediate protein-protein
321 interactions in the assembly and dynamics of the Type-III secretion systems of proteobacteria (84).
322 Using profile-profile searches, we identified a previously uncharacterized protein encoded by genes
323 linked to the yjfJ family of *pspA* genes (*e.g.,* ANH61663.1) as a novel member of the CesT/Tir
324 superfamily (**Fig. 3**; **Tables S1** and **S2**; *e.g.,* ANH61662.1). We also observed that the
325 proteobacterial proteins in the neighborhood of *pspA* (**Fig. 3;** *e.g.,* BAB38581.1), typified by YjfI
326 from *E. coli* (DUF2170; **Table S2**; *e.g.,* BAB38580.1), contained a CesT/Tir superfamily domain (85).
327

328 The dyad encoding CesT/Tir–PspA forms the core of one class of conserved gene neighborhood
329 (**Table S2**). We observed it as either a standalone unit or within larger operons, such as one group
330 combining CesT/Tir–PspA (**Fig. 3**; **Table S3**; ANH61663.1) with i) a gene encoding a membrane-
331 associated protein with the domain architecture TM+Band-7+coiled-coil+flotillin, ii) a novel AAA+-
332 ATPase fused to N-terminal coiled-coil and β-propeller repeats, and iii) a 3TM domain protein

333 prototyped by *E. coli* YqiJ and *B. subtilis* YuaF (previously implicated in resistance to cell-wall
334 targeting antibiotics) (86, 87). In some proteobacteria, this operon also contains a phospholipase D
335 (HKD) superfamily hydrolase (**Fig. 3**). Related abbreviated operons coding only for the Band-
336 7/flotillin-containing protein, the YqiJ/YuaF TM protein, and in some cases, the aforementioned
337 AAA+-ATPase protein is more widely distributed across bacteria and archaea. These might function
338 with PspA homologs encoded elsewhere in the genome.
339
340 The second major group incorporates the dyad into a 6–7 gene operon (**Fig. 3**; **Table S3**; *e.g.,*
341 AAN56746.1) featuring a gene encoding a spermine/spermidine-like synthase domain (86) fused to
342 an N-terminal 7TM transporter-like domain (**Fig. 3**; **Tables S2** and **S3**; *e.g.,* AAN56744.1). This
343 operon encodes three additional uncharacterized proteins, including the YjfL family 8TM protein
344 (DUF350), a novel lipoprotein (Ctha_1186 domain), and a β-strand rich protein that we predict to be
345 intracellularly localized (**Fig. 3**; **Tables S3** and **S4;** DUF4178; *e.g.,* AAN56747.1). In some cases, the
346 last gene in this operon codes for a ribbon-helix-helix (RHH) domain transcription factor (**Fig. 3**;
347 **Tables S2** and **S3**).
348
349 The third group combines the CesT/Tir–PspA dyad (AMJ95269.1) with polyamine metabolism genes
350 encoding an ATP-grasp peptide ligase (AMJ95273.1) related to the glutathionyl spermidine
351 synthetase (**Fig. 3**; **Table S3**). This association was also recently noticed in a study of AdoMet
352 decarboxylase gene linkages (88). The operon also encodes a potassium channel with intracellular-
353 ligand-sensing, a YjfL family 4TM protein (**Tables S2** and **S3**; DUF350), a metal-chelating lipoprotein
354 (DUF1190), and another uncharacterized protein with predicted enzymatic activity (DUF2491).
355 Alternately, two variants of these operons lack the CesT/Tir–PspA dyad: one where its replacement
356 is a gene encoding a distinct protein occurring as either a secreted or lipoprotein version (DUF4247,

357 *e.g.,* CCP45395.1) and a predicted beta-sandwich enzymatic domain with two highly conserved
358 histidines, as well as a conserved aspartic acid and glutamic acid (DUF2617, *e.g.,* CCP45394.1). In
359 the second variant, the dyad is replaced by a protein containing a Band-7 domain fused to C-
360 terminal 2TM and SHOCT domains (see below). These operons encode two further polyamine
361 metabolism genes, an AdoMet decarboxylase, and a polyamine oxidase.
362
363 Lastly, the *Bacillus subtilis ydjFGHL* operon typifies a coupling in Firmicutes of *pspA* (**Table S1**;
364 ANX09535.1) with Band-7 and C-terminal Zinc Ribbon (ZnR) domains in YdjL, and two ZnRs
365 followed by an uncharacterized domain related to YpeB featuring an NTF2-like fold and a TM
366 segment in protein YdjG. The conserved partner of PspA in these systems is the membrane-
367 associated YdjH, which is characterized by a so-called "**TPM phosphatase domain**" followed by
368 a low- complexity segment or a long coiled-coil (**Fig. 3**; **Table S3**) (88). Competing hypotheses
369 suggest that this domain either encodes a generic phosphatase based on its plant homolog (88) or
370 is involved in membrane-associated photosystem II repair and respiratory complex III assembly (89,
371 90). Our finding that it associates with PspA parallels its association with the chaperone CesT/Tir
372 and Band-7 involved in macromolecular assembly. Thus, our result supports the hypothesis that
373 the TPM domain with PspA plays a chaperone-like role in assembly of membrane-linked protein
374 complexes.
375
376 *Membrane dynamics with Chaperone-like domains*
377 The contextual association of PspA with AAA+-ATPase, two distinct families of CesT/Tir type
378 chaperones, Band-7 and TPM domains further supports functional roles of PspA homologs in
379 recruiting macromolecular assembly machinery to the inner leaf of the cell membrane. Polyamines
380 like spermine/spermidine have been implicated in membrane stability in a concentration-dependent
381 manner (91). The repeated coupling of polyamine metabolism genes we identified in PspA-CesT/Tir

382 operons implies that a PspA-based system of membrane dynamics monitors polyamine
383 concentration or aminopropylation of other substrates (92) to alter membrane structure and
384 membrane-associated protein complexes. We propose that these systems link extracellular stress
385 sensing with intracellular membrane dynamics based on the frequent coupling with cell-surface
386 lipoproteins. In particular, flotillin domain-containing proteins are likely to be recruited to lipid
387 subdomains with special components (*e.g.,* cardiolipin-rich inner membrane) to interface with PspA
388 (87).

389

390 *The PspAA domain*

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

400

401 *PspA with PspM (ClgRPspAMN) or Thioredoxin*

402 Certain *pspA* homologs may occur in Cyanobacteria and Actinobacteria as an operonic dyad with
403 a gene encoding an active thioredoxin domain-containing protein (15). PspA in this dyad is typified
404 by corynebacterial RsmP (*e.g.,* AKK09942.1; **Fig. 3**; **Tables S2** and **S3**; **Text S2.3**), which is
405 predominantly found in rod-shaped actinobacteria and regulates cell shape via phosphorylation

406 (93). These PspA homologs either occur with thioredoxin (*e.g.,* AKK09942.1) or with a ClgR-like
407 cHTH and PspM genes (*e.g.,* CCP45543.1; **Fig. 3, Fig. S1**, **Table S3**). Corynebacterium spp. contain
408 two distinct PspA paralogs each displaying one of the above contextual associations ('Paralog' tab,
409 web app). The same thioredoxin family occurs with a different PspA clade (typically two copies) in
410 Cyanobacteria (*e.g.,* AFZ14666.1; Fig. 3; Table S3). The association of the ClgR-PspM and
411 thioredoxin and their mutual exclusion in the operon suggest that they act as alternative regulators
412 of the PspA homologs — repressors in the case of the ClgR, and a potential redox regulator in the
413 case of thioredoxin (**Fig. S1**; **Text S2.3**).
414

415 Operons with two-component systems

416 Two-component systems often associate with the PSP system. For instance, the *liaIHGFSR* operon
417 from *B. subtilis* has been studied in the context of lantibiotic resistance (*e.g.*, ANX06812.1; **Fig. 3**;
418 **Table S3**) (11). In a few paenibacilli, genes encoding an additional PspA and PspC with a C-terminal
419 coiled-coil have been inserted into a *liaIHGFSR*-like operon (*e.g.,* APB74393.1; **Fig. 3**; **Table S3**). In
420 Actinobacteria (*e.g.*, CAB51252.1; **Fig. 3**; **Table S3**), the PspA–PspAA dyad occurs with a firmicute-
421 like two-component system.
422

423 *Classic two-component transcriptional signaling system*

424 We identified two-component signaling system operons linked to either *pspA* or *PspC* in Firmicutes,
425 Actinobacteria, and other clades (**Fig. 3**; web app). In these operons, a membrane-bound histidine
426 kinase (HISKIN) communicates an external stress signal to the response regulator (REC–HTH)
427 protein, triggering a transcriptional response. It is very likely that PspA, LiaI-LiaF-TM, and Toastrack
428 tie into a two-component system regulating membrane integrity and permeability in response to the
429 stress signal. In Actinobacteria, where the HISKIN is fused to PspC, the signal is presumably sensed

430 by PspC. When *pspC/Toastrack* operons with two-component systems lack *pspA* in their
431 immediate operonic neighborhood, they likely recruit PspA proteins from elsewhere in the genome
432 to elicit the stress response function.
433

434 PspA-free variations of domain architectures and gene-neighborhood

435 We next investigated domain architectures and genomic contexts involving PSP components that
436 lack PspA, including those with the PspC and Toastrack domains.
437

438 PspC domain architectures and gene-neighborhood

439 **PspC** is present in most bacteria and the archaeal clades of Euryarchaeota and Asgardarchaeota.
440 Some orthologs of PspC are fused to an N-terminal ZnR domain (*e.g.,* ABC83427.1, **Table S4**),
441 while most other homologs occur either in predicted operons or fusions with multi-TM domains,
442 such as the LiaI-LiaF-TM and PspB (**Fig. 4**; **Table S4**). PspC also fuses to diverse signaling domains
443 such as i) HISKIN from two-component systems (see above); ii) a novel signaling domain we term
444 the HTH-associated α-helical signaling domain (**HAAS)**, which overlaps partly with the Pfam model
445 for DUF1700 (**Table S2**), and iii) the Toastrack domain (**Table S1**). Moreover, actinobacteria contain
446 a Lia-like system without PspA. The HISKIN domain in these systems is fused to N-terminal PspC
447 and LiaF-LiaI-TM domains (*e.g.,* AIJ13865.1) and accompanied by a REC–HTH-containing protein
448 (**Fig. 5**; **Table S4**). This core shares a promoter region with a second gene on the opposite strand
449 containing a PspC domain that is fused to additional TMs and, in some cases, a Toastrack domain
450 (*e.g.,* AIJ13866.1; **Fig. 5**; **Table S4**). These associations strongly imply that PspC acts as a sensor
451 domain to detect and communicate a membrane-proximal signal.
452

453 ## Contextual associations of the Toastrack Domain

454 The **Toastrack** domain repeatedly emerges as a partner to the PSP system. As noted, genes
455 encoding Toastrack-containing protein occur concurrently with *pspA* in several conserved
456 neighborhoods (**Fig. 4**; **Table S4**) and are fused to *pspC* in *pspA*-free settings. We found that the
457 Toastrack domains tend to be C-terminal to various TM domains such as PspC, LiaI-LiaF-TM, and
458 other multi-TM/single-TM domains unique to these systems (**Fig. 4**; **Text S2.4**). These fusions
459 strongly suggest that Toastrack is an intracellular domain. Further, the N-terminal regions of several
460 Toastrack architectures contain at least two divergent homologs of the bi-helical SHOCT domain
461 (*e.g.,* CAB15517.1; **Figs. 4** and **5**; **Tables S1** and **S4**), which we call **SHOCT-like** domains, which
462 are partly detected by the Pfam **DUF1707** model (**Text S2.4**) (94)**.** We observed these SHOCT-like
463 domains at both the N- and C-termini of various globular domains, including Toastrack. Therefore,
464 we predict that this SHOCT-like domain plays a role in anchoring disparate domains to the inner
465 leaf of the membrane.
466
467 The DNA-binding domain **LytTR** may also be encoded by operons carrying Toastrack fused to LiaI-
468 LiaF-TM in Firmicutes (*e.g.,* CAL82154.1; **Fig. 5**; **Table S4**). The Toastrack-containing protein (*e.g.,*
469 CAB15517.1) occurs in an operon with a *pspC* homolog (CAB15516.1) and has been implicated in
470 the activation of the membrane-associated *liaFSR* operon (**Table S1**) and protection against
471 membrane permeabilization (95–97). We also identified *pspA*-free variants of the classic *lia* operon
472 with Toastrack and a two-component system, typified by *vraT* of *Staphylococcus aureus*
473 (ABD31150.1; **Fig. 5**; **Table S4**). These systems carry LiaI-LiaF-TM and Toastrack with a two-
474 component system (*vraSR*) and are involved in methicillin and cell-wall-targeting antibiotic
475 resistance (3, 98–100). A similar organization of LiaI-LiaF-TM and Toastrack containing a protein

476 with a two-component system is found in a few bacterial species (*e.g.,* among Proteobacteria,
477 ABD83157.1, Ignavibacteriae, AFH48155.1; **Table S4**).
478
479 We discovered several other conserved gene neighborhoods centered on Toastrack-containing
480 genes likely to define novel functional analogs of the PSP system with potential roles in membrane-
481 linked stress response. One of these, a four-gene context found across diverse bacterial lineages,
482 includes a previously uncharacterized protein with similarity to the Pfam model DUF2089. The Pfam
483 model **DUF2089** can be divided into an N-terminal ZnR, central HTH, and C-terminal SHOCT-like
484 domains (ADE70705.1), variants of which appear with LiaI-LiaF-TM, B-box, or PspC domains (**Figs.**
485 **4** and **5**; **Tables S2** and **S4**; **Text S2.4**). We propose that these Toastrack contexts function similarly
486 to the classical *lia* operon in transducing membrane-associated signals to a transcriptional output
487 affecting a wide range of genes, including through an operonically linked sigma factor.
488
489 Similarly, we observed operons coupling a membrane-anchored Toastrack-containing protein with
490 an ABC-ATPase, a permease subunit, a **GNTR-HTH** transcription factor with distinct C-terminal α-
491 helical domain (**Fig. 5**; **Tables S2** and **S4**; **Text S2.4**), and a DUF2884 membrane-associated
492 lipoprotein (*e.g.,* AJI31452.1). The latter might function as an extracellular solute-binding partner for
493 the ABC-ATPase and permease components. We found a similar operon in Actinobacteria with the
494 ribbon-helix-helix (**RHH**) domain transcription factor replacing **GNTR-HTH** (**Table S2**; **Text S2.4**).
495 In some Actinobacteria, the Toastrack domain is fused to a SHOCT-like domain or occurs in a
496 transport operon, likely coupling Toastrack to transcriptional regulation, membrane-proximal signal
497 sensing, and transport (**Fig. 5**; **Table S4**).
498

499 Based on these observations, we propose that the two extended β-sheets of the Toastrack domain
500 (formed by the β-helical repeats) facilitate the assembly of sub-membrane signaling complexes by
501 providing two expansive surfaces amenable to protein-protein interactions near the membrane.
502 Protein docking to the Toastrack domain may also activate transcription via fused or associated
503 HTH or LytTR domains, associated two-component systems, or the HAAS-HTH system.
504
505 *The HAAS-PadR-like wHTH systems*
506 We discovered a novel signaling system that frequently co-occurs with PspC and Toastrack
507 domains in gene neighborhoods that do not contain *pspA* (**Figs. 4** and **5**). This system consists of
508 two components, the HAAS domain and the PadR-like wHTH (**Table S2**). The HAAS domain is
509 partly similar to the Yip1 protein family (Pfam models DUF1700 and DUF1129; Pfam clan: Yip1). An
510 HHpred search further unifies this model with the Pfam profile DUF1048 (PDB: 2O3L). We therefore
511 refer to them collectively as the **HAAS** superfamily (HAAS under 'MSA' tab, web app). The core
512 HAAS fold has three consecutive α-helices, with the third helix displaying an unusual kink
513 corresponding to a conserved GxP motif, predicted to form part of a conserved groove that
514 mediates protein-protein interactions. The HAAS domain always occurs in gene-neighborhoods
515 coupled with a protein containing a standalone PadR-like wHTH DNA-binding domain (**Fig. 5**). This
516 co-migration suggests that HAAS transduces a signal from domains fused to it to this wHTH
517 transcription factor.
518
519 When coupled with PspC, the HAAS domain shows three broad themes: 1) it is found as part of a
520 multidomain TM protein with additional PspC, LiaI-LiaF-TM, and Toastrack domains; 2) it is fused
521 directly to a Toastrack domain in a gene neighborhood that also codes for a PspC domain protein,
522 occasionally fused to other TM domains; 3) constitutes a part of a TM protein fused to conserved

523 multi-TM domains other than the LiaI-LiaF-TM or PspC (*e.g.,* VanZ) (**Figs. 4** and **5**; **Table S4**). These
524 gene neighborhoods code for standalone *pspC* genes. Additionally, the HAAS domains occur in
525 PspC-independent contexts, but are typically coupled to the N-terminus of other multi-TM or
526 intracellular sensory domains. We found conserved fusions to at least 5 distinct multi-TM domains
527 (**Figs. 4** and **5**; **Table S4**), such as 1) the FtsX-like TM domains with extracellular solute-binding
528 MacB/PCD domains (*e.g.,* ACO32024.1); 2) FtsW/RodA/SpoVE family of TM domains (*e.g.,*
529 CAC98500.1) (100); 3) various uncharacterized 6TM, 4TM, and 2TM domains. In addition to
530 Toastrack, the HAAS domain may combine with other intracellular domains such as the
531 pentapeptide repeat domains mostly in firmicutes (*e.g.,* AOH56696.1; **Fig. 5**; **Table S4**). The
532 organisms in which these operons occur have Psp components elsewhere in the genome. These
533 proteins might be recruited to the stress response system, as suggested by the effects of PadR-
534 like wHTH deletion studies in the Listeria Lia systems (101).
535
536 In bacteria without the typical *lia* operon (carrying the LiaRS-like two-component system), we noted
537 that the HAAS–PadR-like-wHTH dyad replaces the histidine kinase-receiver domain transcription
538 factor pair. Therefore, we propose that the HAAS-PadR dyad constitutes an alternative to the
539 classical two-component systems. The HAAS domain may undergo a conformational change in
540 response to membrane-linked or intracellular signals detected by the sensor domains (*e.g.,* LiaI-
541 LiaF-TM or PspC) that fuse to or co-occur with the dyad. This conformational change is likely
542 transmitted to the associated PadR-like wHTH via the groove associated with the conserved GxP
543 motif; this change might lead to the release of the transcription factor to mediate a downstream
544 transcriptional response.
545

A unified view of PSP partners and evolution

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

Conclusion

Recent work with the bacterial PSP stress response system suggests that while function is maintained across phyla, the proteins, regulatory mechanisms, and membrane stress response mechanisms vary widely among lineages (12, 15, 26, 36) (12, 15, 25, 32). These variations reflect

570 the utilization of a common peripheral inner membrane protein in lineage-specific envelope
571 dynamics and responses to unique environmental stresses. In this study, we systematically
572 analyzed PSP systems, their phyletic spread, and evolution (**Figs. 1** and **2**), discovering several
573 novel sequence and structural features (*e.g.,* domain architectures) and conserved gene-
574 neighborhoods of PSP components. We identified both classical and newly-predicted partners
575 (**Figs. 2** and **5**) and established their phyletic distributions across the tree of life (**Figs. 2** and **6**). We
576 also established that PspA/Snf7 traces to the LUCA (**Fig. 2**). Indeed, it has been independently
577 shown that the PspA/Vipp1 ESCRT system dates back to the LUCA (19, 20), further strengthening
578 the evolutionary significance of this finding. This confirms that the LUCA possessed a membrane
579 whose curvature and dynamics were mediated by an ancestral PspA/Snf7-like coiled-coil protein
580 assembled into polymeric structures adjacent to the inner-leaf of the membrane. This hypothesis is
581 in line with earlier work showing that LUCA contained some signal-recognition particle GTPases
582 (102). These proteins are unequivocal indicators of the presence of a secretory apparatus
583 translocating proteins across lipid bilayers. Hence, as a corollary they are also indicators of the
584 presence of a membrane. Similarly, the extant components of membrane linked bioenergetics
585 (F1/F0 ATPases or V-Type ATPases) are also indicative of a membrane in the LUCA (103, 104).
586 Thus, we infer that this ancient machinery was robust to the subsequent lineage-specific changes
587 in membrane chemistry, producing different membrane types (*e.g.,* ether-linked vs. ester-linked) in
588 the archaeal, bacterial, and eukaryotic lineages.
589
590 Our novel PSP findings make several new functional predictions: 1) Based on its occurrence with
591 membrane partners (*e.g.,* PspA, LiaI-LiaF-TM, PspC, HAAS, SHOCT-like domains) and transcription
592 regulators or two-component systems (*e.g.,* LiaRS, PadR-like-wHTH), we propose that Toastrack
593 is a membrane-proximal intracellular domain serving as a sub-membrane signaling complex
594 assembly platform; 2) The repeated association of PspA homologs with distinct AAA^{+}-chaperones

595 and a non-ATPase chaperone (*e.g.,* CesT/Tir) or predicted chaperone-like proteins (*e.g.,* Band-7,
596 TPM) suggest that these associations are involved in the assembly of sub- and trans-membrane
597 complexes in response to specific envelope stresses. These observations imply a general role for
598 analogs of ESCRT-like complexes in these bacterial systems; 3) The association of PspA with
599 various transporters and polyamine metabolism systems indicates that the regulation of membrane
600 structure by PspA is associated with changes in solute concentrations and covalent modifications
601 that may affect membrane stability; 4) The discovery of several new versions of the SHOCT-like
602 domain suggests that these domains mediate membrane localization of disparate catalytic and non-
603 catalytic activities, which may subsequently interface with the PSP system; 5) A diverse array of
604 sub-membrane (*e.g.,* Toastrack), TM, and cell-surface domains (*e.g.,* PspC, LiaI-LiaF-TM) co-occur
605 with two-component or HAAS-PadR-like-wHTH systems, hinting at alternatives to conventional
606 two-component signaling that interface with the PSP system in responding to membrane stress.
607
608 Further research into these newly described systems will yield broadly applicable mechanistic
609 insights, given their conservation across swaths of life. For example, SHOCT domains are
610 understudied due to their believed absence in model organisms (94), but our approach identified
611 divergent homologs in the model organism *B. subtilis* (CAB15517.1), which may prompt new
612 studies. The extensive findings from applying gene neighborhood and domain architecture
613 searches support their viability, and such methods help reduce the space of annotated hypothetical
614 proteins without classified domains and add functional clues. Finally, we provide another pillar of
615 support to recent articles' conclusions that PspA/Snf7 homologs were already present in the last
616 universal common ancestor (19, 20), and deepen our understanding of the origins and diversification
617 of all life. All the findings (data, results, and visual summaries) from this work are found in an
618 interactive and queryable web application available at jravilab.shinyapps.io/psp_evolution.
619

620 # Methods

621 ## Query and subject selection

622 We queried classical PSP members — PspA (eight representatives across Proteobacteria,
623 Actinobacteria, Firmicutes (two paralogs), Cyanobacteria, and Archaea); PspM (Rv2743c) and PspN
624 (Rv2742c) (from *M. tuberculosis*); PspB and PspC (from *E. coli*); LiaI, LiaG, and LiaF (from *B. subtilis*)
625 — against all sequenced genomes across the tree of life. We ran homology searches against
626 custom databases of ~6500 completed representative or reference genomes with taxonomic
627 lineages or the NCBI non-redundant database (**Tables S1** and **S2**) (105, 106). All PSP homologs are
628 listed in the web app, with representatives in **Tables S3** and **S4**. We obtained the phyletic order
629 (sequence) from NCBI taxonomy and BV-BRC (105–109).

630

631 ## Identification and characterization of protein homologs

632 We analyzed the classical PSP proteins from the best-studied operons *pspF||ABCDE-G* (from *E.*
633 *coli*), *liaIHGFSR* (from *B. subtilis*), and *clgRpspAMN* (from *M. tuberculosis*) and their phyletic spread
634 to locate and stratify all PSP homologs. We first resolved these PSP proteins into their constituent
635 domains and used each individual domain as a query for all our phylogenetic analyses to ensure an
636 exhaustive search and identification of all related PSP proteins, including remote homologs. This
637 approach allows us to find homologies beyond those from full-length proteins only. We then
638 performed homology searches for each constituent domain across the tree of life (approximately
639 6500 representative or reference genomes). We used a combination of protein domain and
640 orthology databases, iterative homology searches, and multiple sequence alignment to detect
641 domains, signal peptides, and transmembrane (TM) regions to help construct the domain
642 architecture of each query PSP protein. The ubiquitous presence of transcription factor HTH
643 domains and histidine kinase-receiver domain two-component systems in bacterial phyla precluded

644 dedicated searches with these domains; instead, we noted their occurrence in predicted operons
645 alongside core PSP genes to identify transcriptional regulation partners.
646
647 We ensured the identification of a comprehensive set of homologs (close and remote) for each
648 queried protein via iterative searches using PSI-BLAST (110) with sequences of both full-length
649 proteins and the corresponding constituent domains. More distant relationships were determined
650 using profile-profile searches with the HHpred program (111). Protein searches were conducted
651 using homologous copies from multiple species as starting points. We aggregated search results
652 and recorded the numbers of homologs per species and genomes carrying each of the query
653 proteins (web app). These proteins were clustered into orthologous families using the similarity-
654 based clustering program BLASTCLUST (112). SignalP, TMHMM, Phobius, JPred, Pfam, and
655 custom profile databases (113–118) were used to identify signal peptides, TM regions, known
656 domains, and the secondary protein structures in every genome. Homolog information, including
657 domain architectures, is available in the web app ('Data,' 'Domain architectures' tabs).
658

659 ## Neighborhood search

660 Bacterial gene neighborhoods (±7 genes flanking each protein homolog) were retrieved using a Perl
661 script from GenBank (105, 106). Gene orientation, domains, and secondary structures of the
662 neighboring proteins were characterized using the same methods applied to query homologs
663 above. Genomic contexts are available in the web app ('Genomic contexts' tab). We note that
664 eukaryotic components (*e.g.,* Snf7) appearing alone are an artifact of the genomic context analysis
665 being restricted (and relevant) to bacteria and archaea (see web app).
666

Phylogenetic analysis

Multiple sequence alignment (MSA) of the identified homologs was performed using Kalign (119) and MUSCLE (120) with default parameters. Domain/gene trees were constructed with FastTree 2.1 with default parameters (121) and visualized and labels colored in FigTree v1.4.4 (122). In parallel, the PspA/Snf7 MSA was also passed to ModelTest-NG, which reported the best performing substitution model (LG+I+G4+F) (123, 124), and then RAxML-NG-MPI (v1.2.0) (125) on CU's Alpine HPC cluster and a maximum likelihood phylogenetic tree constructed (--all, –model LG+I+G4+F, –tree rand{25} pars{25}, –bs-trees autoMRE, –threads 60). Bootstrapping was performed until RAxML-NG assessed statistical convergence at 480 replicates [**Fig. S2**]. MSA and phylogenetic trees are available in the web app ('Phylogeny' tab).

Network reconstruction

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

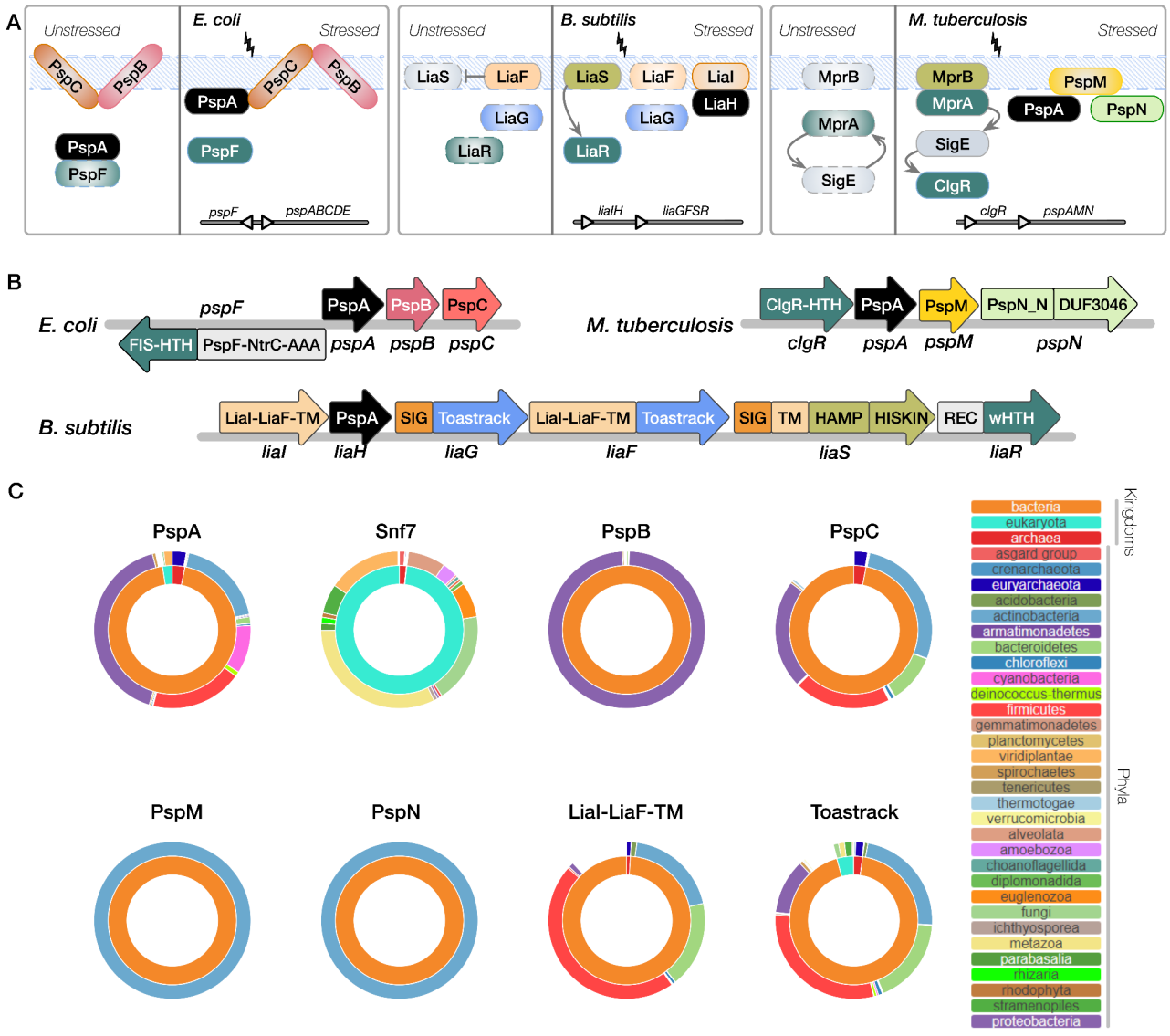
Web application

We built our interactive and searchable web application using R Shiny (126). All data analyses and visualizations were carried out using R or RStudio (127, 128) with R-packages (129–141). Data and results from our study are available on the web application (jravilab.shinyapps.io/psp_evolution). The web app can be queried with the protein accession numbers, domain names, or species cited inline throughout the Results section.

691 # Figures

692 **Figure 1. The phyletic spread of classical Psp members across all major lineages.**

693 **A. The three classical PSP systems** in *E. coli* (*pspF||pspABC*), *M. tuberculosis* (*clgRpspAMN*), and
694 *B. subtilis* (*liaIHGFSR*) are shown. Proteins across three bacterial PSP systems are labeled and
695 colors uniformly (across all figures that show domain architectures) indicate the nature of the
696 protein; ***black***, PspA homolog (*e.g.,* PspA, LiaH); ***teal***, transcription factor/response regulator (*e.g.,*
697 PspF, LiaR, MprA, ClgR) along with partner histidine kinases in ***olive green*** (*e.g.,* LiaS, MprB part
698 of two component systems LiaRS, MprAB); ***orange/yellow***, transmembrane protein (*e.g.,* PspB,
699 PspC, LiaF, LiaI, PspM) embedded in the dashed blue membrane. Arrows represent interaction or
700 activation, or inhibitory feedback in the case of LiaF and LiaS. A simple operon map is shown for
701 each system at the bottom, but is expanded upon in the panel **B**. **B. Domain architectures of the**
702 **classical Psp operons** in *E. coli*, *M. tuberculosis*, and *B. subtilis*. Domains are denoted by
703 rectangular segments inside the block arrow representing each protein labeled below the arrow
704 (*e.g.*, the protein LiaH contains only the PspA domain). The direction of the arrow indicates the
705 direction of transcription. See *Results* for new domain definitions. **C. Phyletic spreads of Psp**
706 **proteins.** Sunburst plots are shown to indicate the lineage distributions (as a fraction of the total
707 number of homologs recorded across lineages) for the homologs of 'classical' domains/protein
708 families of interest: PspA, Snf7, PspB, PspC, PspM, PspN (and DUF3046), LiaI-LiaF-TM, Toastrack.
709 In each plot corresponding to a particular protein, the inner ring corresponds to the proportion of
710 its homologs present in superkingdoms/domains of life, Bacteria, Archaea, and Eukaryota. The
711 outer ring depicts the distribution of the homologs among key phyla. Interactive sunburst plots for
712 each of the PSP proteins are available in the web app. The colors for each lineage (outer and inner
713 rings) are shown separately for **C**.

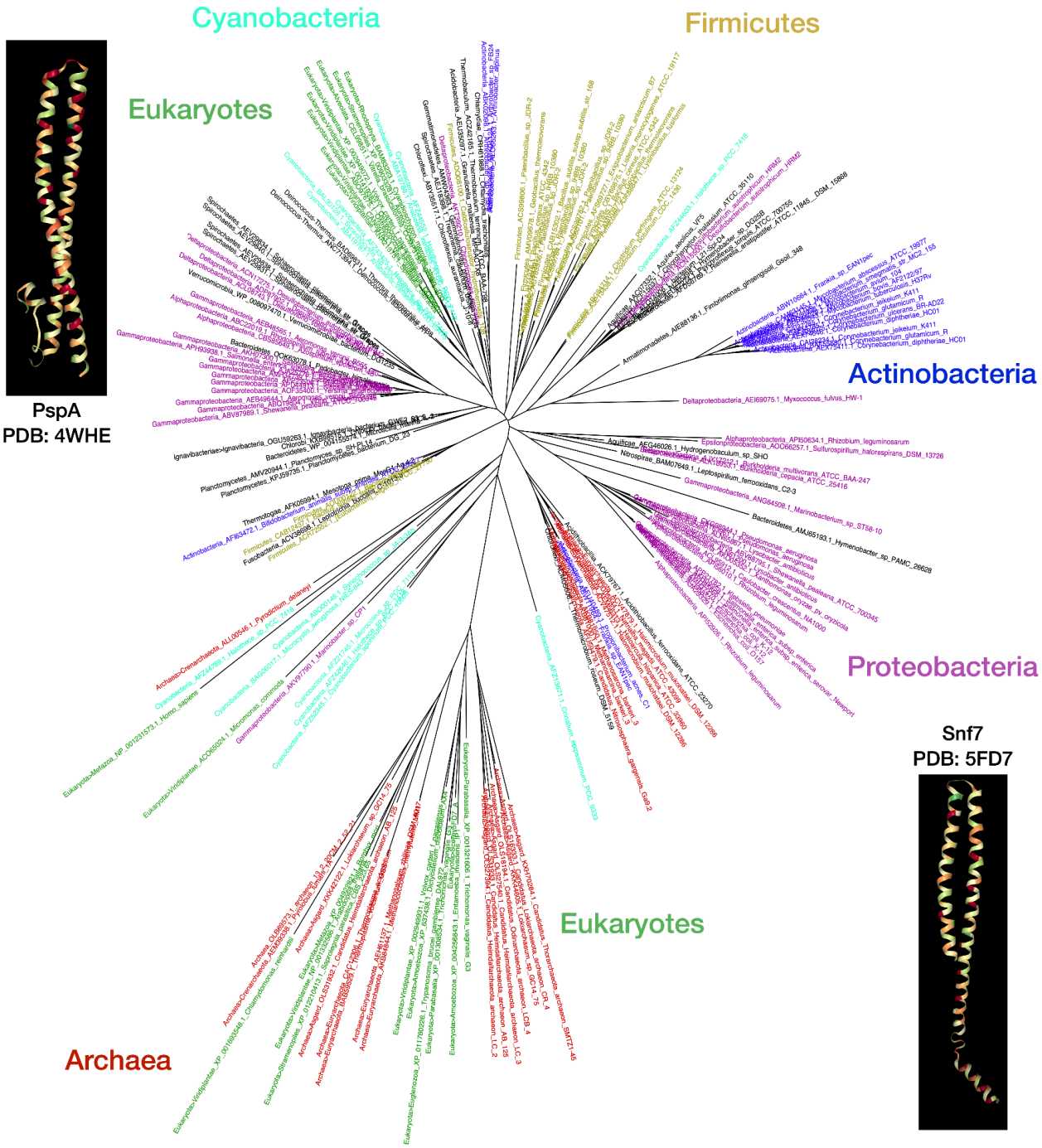

A
Unstressed
E. coli
Stressed
PspC
PspB
PspA
PspF
pspF
pspABCDE
Unstressed
B. subtilis
Stressed
LiaS
LiaF
LiaI
LiaH
LiaG
LiaR
liaIH
liaGFSR
Unstressed
M. tuberculosis
Stressed
MprB
MprA
SigE
PspM
PspN
ClgR
clgR
pspAMN
B
E. coli
pspF
PspA
PspB
PspC
FIS-HTH
PspF-NtrC-AAA
pspA
pspB
pspC
M. tuberculosis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

714

715

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

717 **Phylogenetic tree of PspA homologs across the tree of life.** The phylogenetic tree was
718 constructed using FastTree and visualized with FigTree (parallel RAxML-NG generated PspA tree
719 with confidence values is shown in **Fig. S2**) from a multiple sequence alignment (using kalign3) of
720 representative PspA homologs across all major superkingdoms and phyla (see *Methods*; web app).
721 The key lineages are labeled next to distinct clusters of similar PspA proteins; leaf colors match the
722 lineage labels. The insets show the 3D structures for PspA (4WHE) and Snf7 (5FD7) from the Protein
723 Data Bank. The tree leaves are labeled by lineage, species, and accession numbers. A text-
724 searchable, scaled vector graphics version of this tree is available through the web app under the
725 Phylogeny tab. Representatives for PspA/Snf7 homologs are also available in **Table S3**. As a gene
726 tree, most homologs do cluster among evolutionarily similar species, though evidence of
727 widespread horizontal gene transfer is also evident. Two clusters of eukaryotic genes appear: Snf7
728 in Eukaryota and Asgardarchaeota group, mirroring recent research suggesting eukaryotic life arose
729 from this clade; and Vipp1 among Eukaryota and Cyanobacteria group, again reflecting the shared
730 origin of the protein in plants from Cyanobacteria.


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

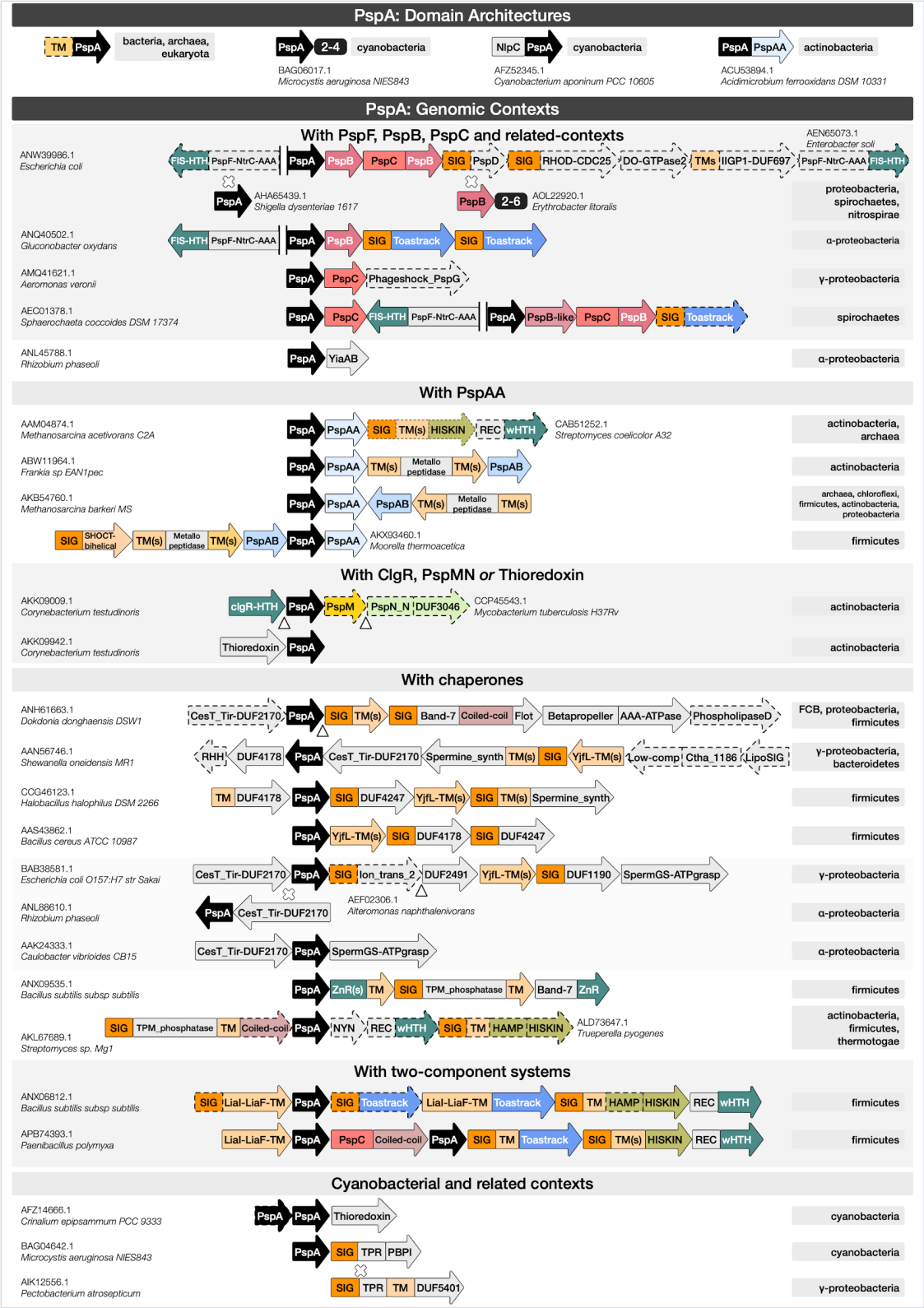
731

732

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

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

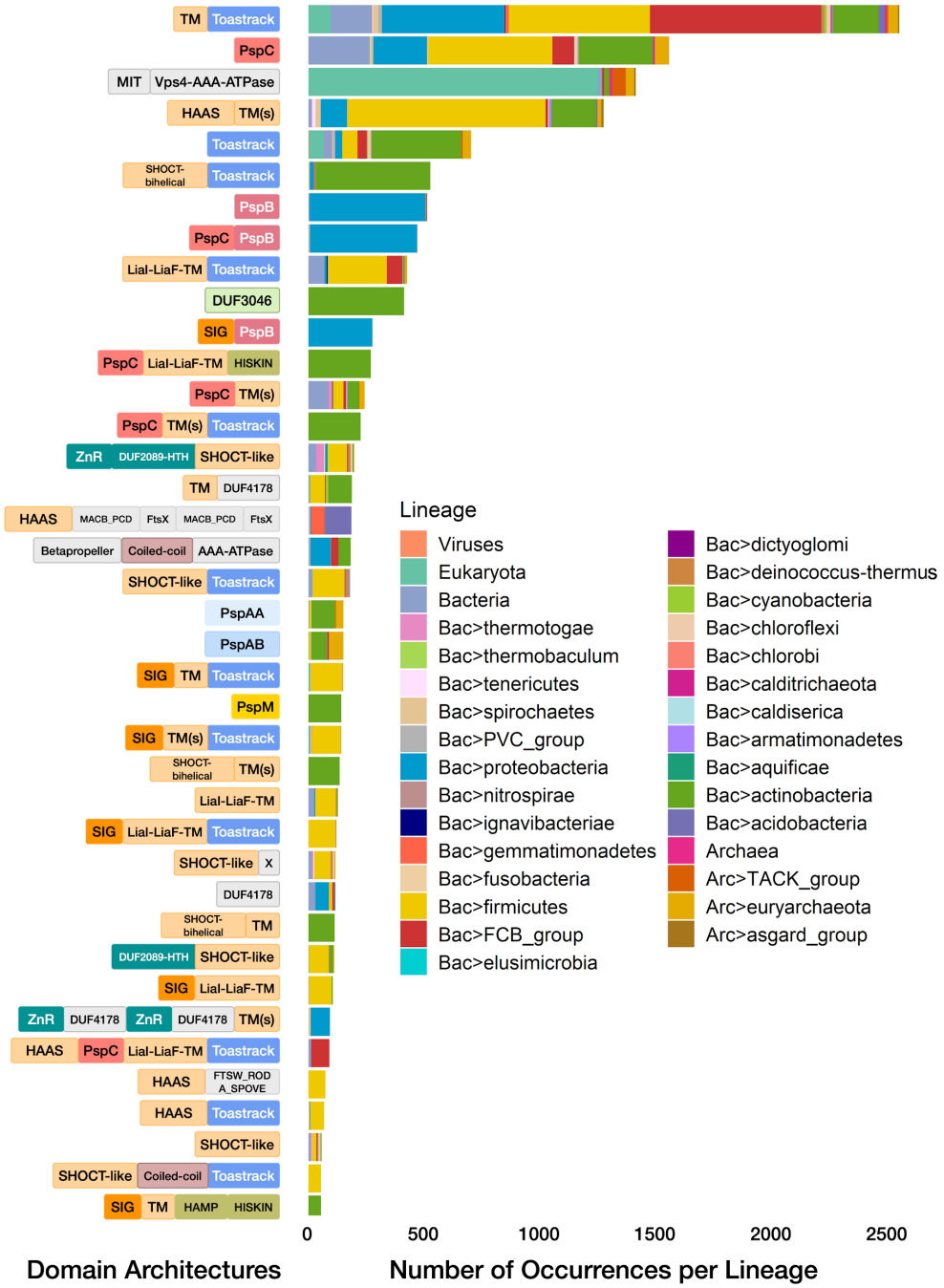
architectures, genomic contexts, and lineages are shown in the web app (under the 'Data,' 'Domain architectures,' and 'Genomic contexts' tabs).


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

758

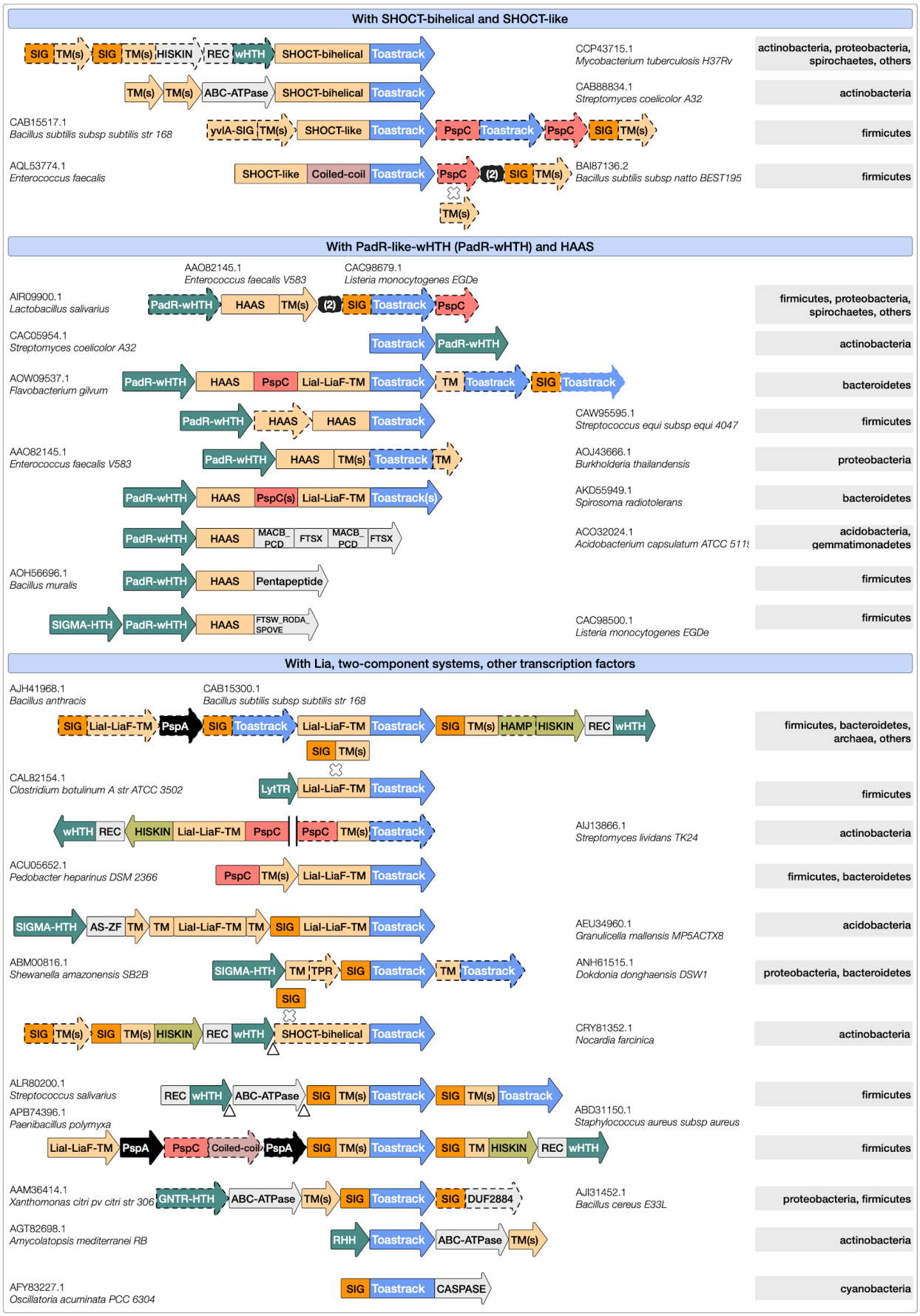
759 **Figure 4. Lineage spread of PspA-free domain architectures.**

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


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

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

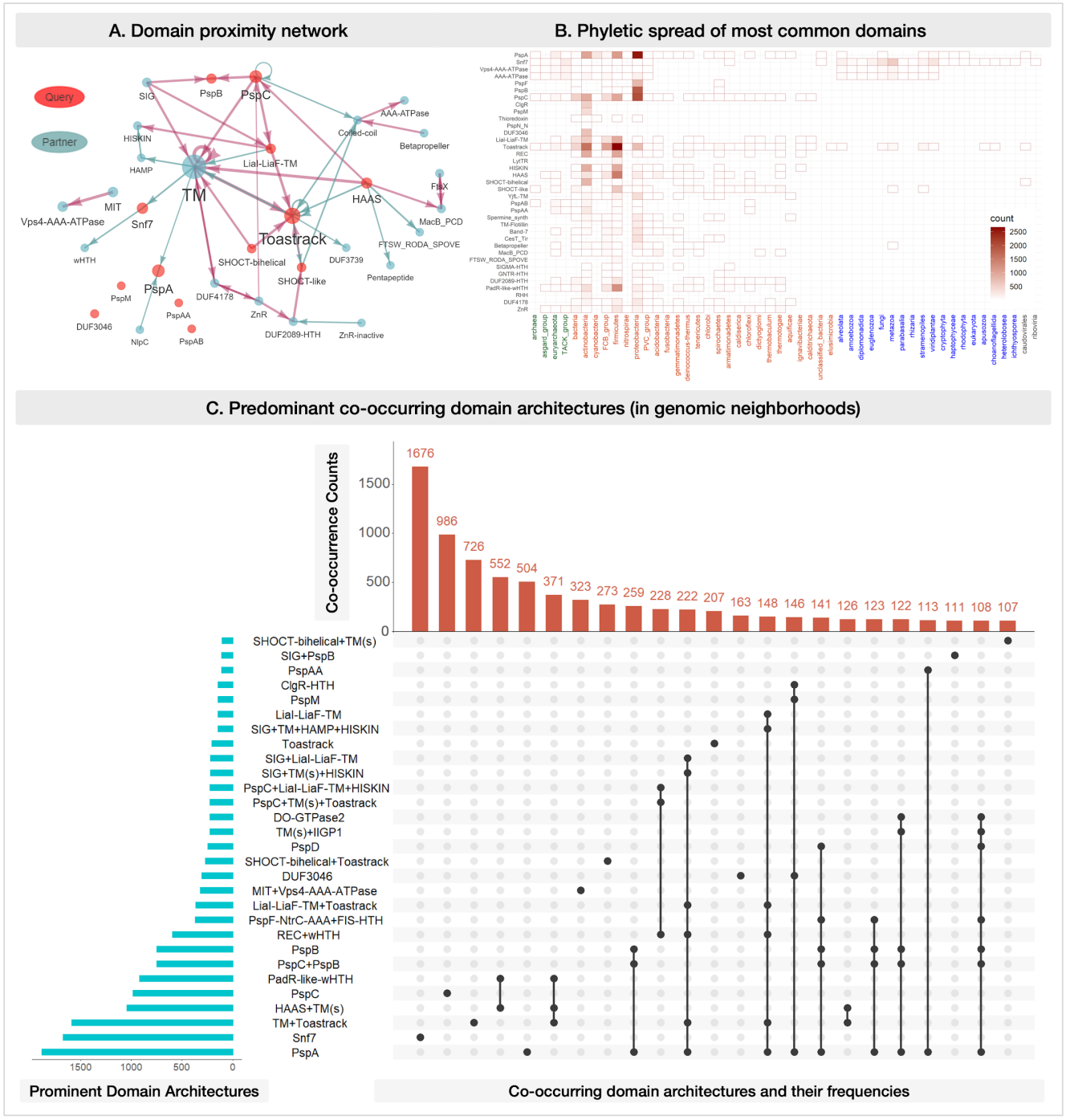
772 The genomic contexts and key lineage memberships are presented using the same schematic as
773 in **Figure 3**. The focus is on PSP partner domains such as Toastrack (blue), PspC, LiaI-LiaF-TM,
774 HAAS, SHOCT-bi-helical (in shades of orange since they are transmembrane domains), and the
775 various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


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

780 **Figure 6. PSP consolidated.**

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

domain architectures and genomic contexts for the homologs of all Psp member proteins are available in the [web app](web app).



Declarations

Acknowledgments

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

Funding

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

Author Contributions

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

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

Data Availability and Reuse

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

840 # References

841 1. van den Brink-van der Laan E, Killian JA, de Kruijff B. 2004. Nonbilayer lipids affect peripheral
842 and integral membrane proteins via changes in the lateral pressure profile. Biochim Biophys
843 Acta 1666:275–288.
844 2. Kobayashi R, Suzuki T, Yoshida M. 2007. Escherichia coli phage-shock protein A (PspA) binds
845 to membrane phospholipids and repairs proton leakage of the damaged membranes. Mol
846 Microbiol 66:100–109.
847 3. Jordan S, Hutchings MI, Mascher T. 2008. Cell envelope stress response in Gram-positive
848 bacteria. FEMS Microbiol Rev 32:107–146.
849 4. Raivio TL. 2005. Envelope stress responses and Gram-negative bacterial pathogenesis. Mol
850 Microbiol 56:1119–1128.
851 5. Rowley G, Spector M, Kormanec J, Roberts M. 2006. Pushing the envelope: extracytoplasmic
852 stress responses in bacterial pathogens. Nat Rev Microbiol 4:383–394.
853 6. Omardien S, Drijfhout JW, van Veen H, Schachtschabel S, Riool M, Hamoen LW, Brul S, Zaat
854 SAJ. 2018. Synthetic antimicrobial peptides delocalize membrane bound proteins thereby
855 inducing a cell envelope stress response. Biochim Biophys Acta Biomembr 1860:2416–2427.
856 7. Wenzel M, Kohl B, Münch D, Raatschen N, Albada HB, Hamoen L, Metzler-Nolte N, Sahl H-G,
857 Bandow JE. 2012. Proteomic response of Bacillus subtilis to lantibiotics reflects differences in
858 interaction with the cytoplasmic membrane. Antimicrob Agents Chemother 56:5749–5757.
859 8. Guest RL, Wang J, Wong JL, Raivio TL. 2017. A Bacterial Stress Response Regulates
860 Respiratory Protein Complexes To Control Envelope Stress Adaptation. J Bacteriol 199.
861 9. Gottesman S. 2017. Stress Reduction, Bacterial Style. J Bacteriol 199.
862 10. Brissette JL, Weiner L, Ripmaster TL, Model P. 1991. Characterization and sequence of the
863 Escherichia coli stress-induced psp operon. J Mol Biol 220:35–48.
864 11. Mascher T, Zimmer SL, Smith T-A, Helmann JD. 2004. Antibiotic-inducible promoter regulated
865 by the cell envelope stress-sensing two-component system LiaRS of Bacillus subtilis.
866 Antimicrob Agents Chemother 48:2888–2896.
867 12. Huvet M, Toni T, Sheng X, Thorne T, Jovanovic G, Engl C, Buck M, Pinney JW, Stumpf MPH.
868 2011. The evolution of the phage shock protein response system: interplay between protein
869 function, genomic organization, and system function. Mol Biol Evol 28:1141–1155.
870 13. McDonald C, Jovanovic G, Ces O, Buck M. 2015. Membrane Stored Curvature Elastic Stress
871 Modulates Recruitment of Maintenance Proteins PspA and Vipp1. mBio 6:e01188-01115.
872 14. Datta P, Ravi J, Guerrini V, Chauhan R, Neiditch MB, Shell SS, Fortune SM, Hancioglu B,
873 Igoshin OA, Gennaro ML. 2015. The Psp system of Mycobacterium tuberculosis integrates
874 envelope stress-sensing and envelope-preserving functions. Mol Microbiol 97:408–422.
875 15. Ravi J, Anantharaman V, Aravind L, Gennaro ML. 2018. Variations on a theme: evolution of the
876 phage-shock-protein system in Actinobacteria. Antonie Van Leeuwenhoek 111:753–760.
877 16. Heidrich J, Thurotte A, Schneider D. 2017. Specific interaction of IM30/Vipp1 with
878 cyanobacterial and chloroplast membranes results in membrane remodeling and eventually in
879 membrane fusion. Biochim Biophys Acta Biomembr 1859:537–549.
880 17. McDonald C, Jovanovic G, Wallace BA, Ces O, Buck M. 2017. Structure and function of PspA
881 and Vipp1 N-terminal peptides: Insights into the membrane stress sensing and mitigation.
882 Biochim Biophys Acta Biomembr 1859:28–39.

883 18. Lee I-H, Kai H, Carlson L-A, Groves JT, Hurley JH. 2015. Negative membrane curvature
884 catalyzes nucleation of endosomal sorting complex required for transport (ESCRT)-III assembly.
885 Proc Natl Acad Sci U S A 112:15892–15897.
886 19. Liu J, Tassinari M, Souza DP, Naskar S, Noel JK, Bohuszewicz O, Buck M, Williams TA, Baum
887 B, Low HH. 2021. Bacterial Vipp1 and PspA are members of the ancient ESCRT-III membrane-
888 remodeling superfamily. Cell 184:3660-3673.e18.
889 20. Junglas B, Huber ST, Heidler T, Schlösser L, Mann D, Hennig R, Clarke M, Hellmann N,
890 Schneider D, Sachse C. 2021. PspA adopts an ESCRT-III-like fold and remodels bacterial
891 membranes. Cell 184:3674-3688.e18.
892 21. Horinouchi T, Suzuki S, Hirasawa T, Ono N, Yomo T, Shimizu H, Furusawa C. 2015. Phenotypic
893 convergence in bacterial adaptive evolution to ethanol stress. BMC Evol Biol 15:180.
894 22. Furusawa C, Irie N. 2020. Toward understanding of evolutionary constraints: experimental and
895 theoretical approaches. Biophys Rev 12:1155–1161.
896 23. Kleerebezem M, Crielaard W, Tommassen J. 1996. Involvement of stress protein PspA (phage
897 shock protein A) of Escherichia coli in maintenance of the protonmotive force under stress
898 conditions. EMBO J 15:162–171.
899 24. Huvet M, Toni T, Tan H, Jovanovic G, Engl C, Buck M, Stumpf MPH. 2009. Model-based
900 evolutionary analysis: the natural history of phage-shock stress response. Biochem Soc Trans
901 37:762–767.
902 25. Vothknecht UC, Otters S, Hennig R, Schneider D. 2012. Vipp1: a very important protein in
903 plastids?! J Exp Bot 63:1699–1712.
904 26. Bultema JB, Fuhrmann E, Boekema EJ, Schneider D. 2010. Vipp1 and PspA: Related but not
905 twins. Commun Integr Biol 3:162–165.
906 27. Westphal S, Heins L, Soll J, Vothknecht UC. 2001. Vipp1 deletion mutant of Synechocystis: a
907 connection between bacterial phage shock and thylakoid biogenesis? Proc Natl Acad Sci U S A
908 98:4243–4248.
909 28. Heidrich J, Wulf V, Hennig R, Saur M, Markl J, Sönnichsen C, Schneider D. 2016. Organization
910 into Higher Ordered Ring Structures Counteracts Membrane Binding of IM30, a Protein
911 Associated with Inner Membranes in Chloroplasts and Cyanobacteria. J Biol Chem 291:14954–
912 14962.
913 29. Theis J, Gupta TK, Klingler J, Wan W, Albert S, Keller S, Engel BD, Schroda M. 2019. VIPP1
914 rods engulf membranes containing phosphatidylinositol phosphates. 1. Sci Rep 9:8725.
915 30. Zhang L, Kato Y, Otters S, Vothknecht UC, Sakamoto W. 2012. Essential role of VIPP1 in
916 chloroplast envelope maintenance in Arabidopsis. Plant Cell 24:3695–3707.
917 31. Armstrong RM, Adams KL, Zilisch JE, Bretl DJ, Sato H, Anderson DM, Zahrt TC. 2016. Rv2744c
918 Is a PspA Ortholog That Regulates Lipid Droplet Homeostasis and Nonreplicating Persistence in
919 Mycobacterium tuberculosis. J Bacteriol 198:1645–1661.
920 32. Gupta TK, Klumpe S, Gries K, Heinz S, Wietrzynski W, Ohnishi N, Niemeyer J, Spaniol B,
921 Schaffer M, Rast A, Ostermeier M, Strauss M, Plitzko JM, Baumeister W, Rudack T, Sakamoto
922 W, Nickelsen J, Schuller JM, Schroda M, Engel BD. 2021. Structural basis for VIPP1
923 oligomerization and maintenance of thylakoid membrane integrity. Cell 184:3643-3659.e23.
924 33. Joly N, Engl C, Jovanovic G, Huvet M, Toni T, Sheng X, Stumpf MPH, Buck M. 2010. Managing
925 membrane stress: the phage shock protein (Psp) response, from molecular mechanisms to
926 physiology. FEMS Microbiol Rev 34:797–827.

927 34. Eiamphungporn W, Helmann JD. 2008. The Bacillus subtilis sigma(M) regulon and its
928 contribution to cell envelope stress responses. Mol Microbiol 67:830–848.
929 35. Jordan S, Junker A, Helmann JD, Mascher T. 2006. Regulation of LiaRS-dependent gene
930 expression in bacillus subtilis: identification of inhibitor proteins, regulator binding sites, and
931 target genes of a conserved cell envelope stress-sensing two-component system. J Bacteriol
932 188:5153–5166.
933 36. Manganelli R, Gennaro ML. 2017. Protecting from Envelope Stress: Variations on the Phage-
934 Shock-Protein Theme. Trends Microbiol 25:205–216.
935 37. Popp PF, Gumerov VM, Andrianova EP, Bewersdorf L, Mascher T, Zhulin IB, Wolf D. 2022.
936 Phyletic Distribution and Diversification of the Phage Shock Protein Stress Response System in
937 Bacteria and Archaea. mSystems 7:e01348-21.
938 38. van der Laan M, Urbanus ML, Ten Hagen-Jongman CM, Nouwen N, Oudega B, Harms N,
939 Driessen AJM, Luirink J. 2003. A conserved function of YidC in the biogenesis of respiratory
940 chain complexes. Proc Natl Acad Sci U S A 100:5801–5806.
941 39. Maxson ME, Darwin AJ. 2006. Multiple promoters control expression of the Yersinia
942 enterocolitica phage-shock-protein A (pspA) operon. Microbiol Read Engl 152:1001–1010.
943 40. Weiner L, Model P. 1994. Role of an Escherichia coli stress-response operon in stationary-
944 phase survival. Proc Natl Acad Sci U S A 91:2191–2195.
945 41. Jovanovic G, Lloyd LJ, Stumpf MPH, Mayhew AJ, Buck M. 2006. Induction and function of the
946 phage shock protein extracytoplasmic stress response in Escherichia coli. J Biol Chem
947 281:21147–21161.
948 42. Mitchell AM, Silhavy TJ. 2019. Envelope stress responses: balancing damage repair and
949 toxicity. Nat Rev Microbiol 17:417–428.
950 43. Engl C, Beek AT, Bekker M, de Mattos JT, Jovanovic G, Buck M. 2011. Dissipation of proton
951 motive force is not sufficient to induce the phage shock protein response in Escherichia coli.
952 Curr Microbiol 62:1374–1385.
953 44. Weiner L, Brissette JL, Model P. 1991. Stress-induced expression of the Escherichia coli phage
954 shock protein operon is dependent on sigma 54 and modulated by positive and negative
955 feedback mechanisms. Genes Dev 5:1912–1923.
956 45. Jovanovic G, Engl C, Buck M. 2009. Physical, functional and conditional interactions between
957 ArcAB and phage shock proteins upon secretin-induced stress in Escherichia coli. Mol
958 Microbiol 74:16–28.
959 46. Brissette JL, Russel M, Weiner L, Model P. 1990. Phage shock protein, a stress protein of
960 Escherichia coli. Proc Natl Acad Sci U S A 87:862–866.
961 47. Yamaguchi S, Reid DA, Rothenberg E, Darwin AJ. 2013. Changes in Psp protein binding
962 partners, localization and behaviour upon activation of the Yersinia enterocolitica phage shock
963 protein response. Mol Microbiol 87:656–671.
964 48. Wolf D, Kalamorz F, Wecke T, Juszczak A, Mäder U, Homuth G, Jordan S, Kirstein J, Hoppert
965 M, Voigt B, Hecker M, Mascher T. 2010. In-depth profiling of the LiaR response of Bacillus
966 subtilis. J Bacteriol 192:4680–4693.
967 49. Jordan S, Rietkötter E, Strauch MA, Kalamorz F, Butcher BG, Helmann JD, Mascher T. 2007.
968 LiaRS-dependent gene expression is embedded in transition state regulation in Bacillus subtilis.
969 Microbiol Read Engl 153:2530–2540.
970 50. Castelle CJ, Banfield JF. 2018. Major New Microbial Groups Expand Diversity and Alter our
971 Understanding of the Tree of Life. Cell 172:1181–1197.

51. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nat Microbiol 1:16048.
52. Flores-Kim J, Darwin AJ. 2016. The Phage Shock Protein Response. Annu Rev Microbiol 70:83–101.
53. Darwin AJ, Miller VL. 2001. The psp locus of Yersinia enterocolitica is required for virulence and for growth in vitro when the Ysc type III secretion system is produced. Mol Microbiol 39:429–444.
54. Becker LA, Bang I-S, Crouch M-L, Fang FC. 2005. Compensatory role of PspA, a member of the phage shock protein operon, in rpoE mutant Salmonella enterica serovar Typhimurium. Mol Microbiol 56:1004–1016.
55. Elderkin S, Bordes P, Jones S, Rappas M, Buck M. 2005. Molecular determinants for PspA-mediated repression of the AAA transcriptional activator PspF. J Bacteriol 187:3238–3248.
56. Suntharalingam P, Senadheera MD, Mair RW, Lévesque CM, Cvitkovitch DG. 2009. The LiaFSR system regulates the cell envelope stress response in Streptococcus mutans. J Bacteriol 191:2973–2984.
57. Aravind L, Ponting CP. 1999. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. FEMS Microbiol Lett 176:111–116.
58. Maxson ME, Darwin AJ. 2006. PspB and PspC of Yersinia enterocolitica are dual function proteins: regulators and effectors of the phage-shock-protein response. Mol Microbiol 59:1610–1623.
59. Gueguen E, Savitzky DC, Darwin AJ. 2009. Analysis of the Yersinia enterocolitica PspBC proteins defines functional domains, essential amino acids and new roles within the phage-shock-protein response. Mol Microbiol 74:619–633.
60. Wallrodt I, Jelsbak L, Thomsen LE, Brix L, Lemire S, Gautier L, Nielsen DS, Jovanovic G, Buck M, Olsen JE. 2014. Removal of the phage-shock protein PspB causes reduction of virulence in Salmonella enterica serovar Typhimurium independently of NRAMP1. J Med Microbiol 63:788–795.
61. Kleine B, Chattopadhyay A, Polen T, Pinto D, Mascher T, Bott M, Brocker M, Freudl R. 2017. The three-component system EsrISR regulates a cell envelope stress response in Corynebacterium glutamicum. Mol Microbiol 106:719–741.
62. Srivastava D, Moumene A, Flores-Kim J, Darwin AJ. 2017. Psp Stress Response Proteins Form a Complex with Mislocalized Secretins in the Yersinia enterocolitica Cytoplasmic Membrane. mBio 8.
63. Flores-Kim J, Darwin AJ. 2012. Phage shock protein C (PspC) of Yersinia enterocolitica is a polytopic membrane protein with implications for regulation of the Psp stress response. J Bacteriol 194:6548–6559.
64. Datta P, Ravi J, Guerrini V, Chauhan R, Neiditch MB, Shell SS, Fortune SM, Hancioglu B, Igoshin OA, Gennaro ML. 2015. The Psp system of Mycobacterium tuberculosis integrates envelope stress-sensing and envelope-preserving functions. Mol Microbiol 97:408–422.
65. Burki F, Shalchian-Tabrizi K, Pawlowski J. 2008. Phylogenomics reveals a new “megagroup” including most photosynthetic eukaryotes. Biol Lett 4:366–369.

1015 66. Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AJ. 2009. Phylogenomic
1016 analyses support the monophyly of Excavata and resolve relationships among eukaryotic
1017 "supergroups." Proc Natl Acad Sci U S A 106:3859–3864.
1018 67. Caillat C, Maity S, Miguet N, Roos WH, Weissenhorn W. 2019. The role of VPS4 in ESCRT-III
1019 polymer remodeling. Biochem Soc Trans 47:441–448.
1020 68. Schöneberg J, Pavlin MR, Yan S, Righini M, Lee I-H, Carlson L-A, Bahrami AH, Goldman DH,
1021 Ren X, Hummer G, Bustamante C, Hurley JH. 2018. ATP-dependent force generation and
1022 membrane scission by ESCRT-III and Vps4. Science 362:1423–1428.
1023 69. Weiss P, Huppert S, Kölling R. 2009. Analysis of the dual function of the ESCRT-III protein Snf7
1024 in endocytic trafficking and in gene expression. Biochem J 424:89–97.
1025 70. Godinho RM da C, Crestani J, Kmetzsch L, Araujo G de S, Frases S, Staats CC, Schrank A,
1026 Vainstein MH, Rodrigues ML. 2014. The vacuolar-sorting protein Snf7 is required for export of
1027 virulence determinants in members of the Cryptococcus neoformans complex. Sci Rep 4:6198.
1028 71. Webster BM, Colombi P, Jäger J, Lusk CP. 2014. Surveillance of nuclear pore complex
1029 assembly by ESCRT-III/Vps4. Cell 159:388–401.
1030 72. Zhang L, Kondo H, Kamikubo H, Kataoka M, Sakamoto W. 2016. VIPP1 Has a Disordered C-
1031 Terminal Tail Necessary for Protecting Photosynthetic Membranes against Stress. Plant Physiol
1032 171:1983–1995.
1033 73. Anantharaman V, Aravind L. 2003. Evolutionary history, structural features and biochemical
1034 diversity of the NlpC/P60 superfamily of enzymes. Genome Biol 4:R11.
1035 74. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez
1036 C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST. 2002. A
1037 new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S
1038 A 99:3684–3689.
1039 75. Tang S, Henne WM, Borbat PP, Buchkovich NJ, Freed JH, Mao Y, Fromme JC, Emr SD. 2015.
1040 Structural basis for activation, assembly and membrane binding of ESCRT-III Snf7 filaments.
1041 eLife 4.
1042 76. Anantharaman V, Koonin EV, Aravind L. 2001. Regulatory potential, phyletic distribution and
1043 evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol 307:1271–1292.
1044 77. Kaur G, Burroughs AM, Iyer LM, Aravind L. 2020. Highly regulated, diversifying NTP-dependent
1045 biological conflict systems with implications for the emergence of multicellularity. eLife 9.
1046 78. Tsai JT, Sung N, Lee J, Chang C, Lee S, Tsai FTF. 2019. Crystal Structure of the YcjX Stress
1047 Protein Reveals a Ras-Like GTP-Binding Protein. J Mol Biol 431:3179–3190.
1048 79. Wollert T, Wunder C, Lippincott-Schwartz J, Hurley JH. 2009. Membrane scission by the
1049 ESCRT-III complex. Nature 458:172–177.
1050 80. Frohn BP, Härtel T, Cox J, Schwille P. 2022. Tracing back variations in archaeal ESCRT-based
1051 cell division to protein domain architectures. PloS One 17:e0266395.
1052 81. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, Seitz
1053 KW, Anantharaman K, Starnawski P, Kjeldsen KU, Stott MB, Nunoura T, Banfield JF, Schramm
1054 A, Baker BJ, Spang A, Ettema TJG. 2017. Asgard archaea illuminate the origin of eukaryotic
1055 cellular complexity. Nature 541:353–358.
1056 82. Burroughs AM, Aravind L. 2016. RNA damage in biological conflicts and the diversity of
1057 responding RNA repair systems. Nucleic Acids Res 44:8525–8555.
1058 83. Gehl B, Sweetlove LJ. 2014. Mitochondrial Band-7 family proteins: scaffolds for respiratory
1059 chain assembly? Front Plant Sci 5:141.

1060 84. Delahay RM, Shaw RK, Elliott SJ, Kaper JB, Knutton S, Frankel G. 2002. Functional analysis of
1061 the enteropathogenic Escherichia coli type III secretion system chaperone CesT identifies
1062 domains that mediate substrate interactions. Mol Microbiol 43:61–73.
1063 85. Jovanovic G, Mehta P, McDonald C, Davidson AC, Uzdavinys P, Ying L, Buck M. 2014. The N-
1064 terminal amphipathic helices determine regulatory and effector functions of phage shock protein
1065 A (PspA) in Escherichia coli. J Mol Biol 426:1498–1511.
1066 86. Lee M-J, Huang C-Y, Sun Y-J, Huang H. 2005. Cloning and characterization of spermidine
1067 synthase and its implication in polyamine biosynthesis in Helicobacter pylori strain 26695.
1068 Protein Expr Purif 43:140–148.
1069 87. Jovanovic G, Mehta P, Ying L, Buck M. 2014. Anionic lipids and the cytoskeletal proteins MreB
1070 and RodZ define the spatio-temporal distribution and function of membrane stress controller
1071 PspA in Escherichia coli. Microbiol Read Engl 160:2374–2386.
1072 88. Cerutti ML, Otero LH, Smal C, Pellizza L, Goldbaum FA, Klinke S, Aran M. 2017. Structural and
1073 functional characterization of a cold adapted TPM-domain with ATPase/ADPase activity. J
1074 Struct Biol 197:201–209.
1075 89. Wegener KM, Bennewitz S, Oelmüller R, Pakrasi HB. 2011. The Psb32 protein aids in repairing
1076 photodamaged photosystem II in the cyanobacterium Synechocystis 6803. Mol Plant 4:1052–
1077 1061.
1078 90. Sirpiö S, Allahverdiyeva Y, Suorsa M, Paakkarinen V, Vainonen J, Battchikova N, Aro E-M. 2007.
1079 TLP18.3, a novel thylakoid lumen protein regulating photosystem II repair cycle. Biochem J
1080 406:415–425.
1081 91. Ballas SK, Mohandas N, Marton LJ, Shohet SB. 1983. Stabilization of erythrocyte membranes
1082 by polyamines. Proc Natl Acad Sci U S A 80:1942–1946.
1083 92. Li B, Kurihara S, Kim SH, Liang J, Michael AJ. 2019. A polyamine-independent role for S-
1084 adenosylmethionine decarboxylase. Biochem J 476:2579–2594.
1085 93. Fiuza M, Letek M, Leiba J, Villadangos AF, Vaquera J, Zanella-Cléon I, Mateos LM, Molle V, Gil
1086 JA. 2010. Phosphorylation of a novel cytoskeletal protein (RsmP) regulates rod-shaped
1087 morphology in Corynebacterium glutamicum. J Biol Chem 285:29387–29397.
1088 94. Eberhardt RY, Bartholdson SJ, Punta M, Bateman A. 2013. The SHOCT domain: a widespread
1089 domain under-represented in model organisms. PloS One 8:e57848.
1090 95. Kingston AW, Liao X, Helmann JD. 2013. Contributions of the σ(W) , σ(M) and σ(X) regulons to
1091 the lantibiotic resistome of Bacillus subtilis. Mol Microbiol 90:502–518.
1092 96. Miller C, Kong J, Tran TT, Arias CA, Saxer G, Shamoo Y. 2013. Adaptation of Enterococcus
1093 faecalis to daptomycin reveals an ordered progression to resistance. Antimicrob Agents
1094 Chemother 57:5373–5383.
1095 97. Wang L, Xiao S, Chen X, Chen S, Wang S, Wang C, Tang Y, Dong F. 2019. ytiB and ythA Genes
1096 Reduce the Uranium Removal Capacity of Bacillus atrophaeus. Int J Mol Sci 20.
1097 98. Boyle-Vavra S, Yin S, Jo DS, Montgomery CP, Daum RS. 2013. VraT/YvqF is required for
1098 methicillin resistance and activation of the VraSR regulon in Staphylococcus aureus. Antimicrob
1099 Agents Chemother 57:83–95.
1100 99. Gardete S, Wu SW, Gill S, Tomasz A. 2006. Role of VraSR in antibiotic resistance and antibiotic-
1101 induced stress response in Staphylococcus aureus. Antimicrob Agents Chemother 50:3424–
1102 3434.
1103 100. Wolf D, Domínguez-Cuevas P, Daniel RA, Mascher T. 2012. Cell envelope stress response in
1104 cell wall-deficient L-forms of Bacillus subtilis. Antimicrob Agents Chemother 56:5907–5915.

1105 101. Hauf S, Möller L, Fuchs S, Halbedel S. 2019. PadR-type repressors controlling production of a
1106 non-canonical FtsW/RodA homologue and other trans-membrane proteins. Sci Rep 9:10023.
1107 102. Leipe DD, Wolf YI, Koonin EV, Aravind L. 2002. Classification and evolution of P-loop GTPases
1108 and related ATPases. J Mol Biol 317:41–72.
1109 103. Lane N, Martin WF. 2012. The origin of membrane bioenergetics. Cell 151:1406–1416.
1110 104. Sojo V, Pomiankowski A, Lane N. 2014. A bioenergetic basis for membrane divergence in
1111 archaea and bacteria. PLoS Biol 12:e1001926.
1112 105. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B,
1113 Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin
1114 V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W,
1115 Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K,
1116 Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-
1117 Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A,
1118 Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. 2016. Reference sequence (RefSeq)
1119 database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic
1120 Acids Res 44:D733-745.
1121 106. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2016. GenBank. Nucleic Acids Res
1122 44:D67–D72.
1123 107. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, Dempsey DM, Dickerman A,
1124 Dietrich EM, Kenyon RW, Kuscuoglu M, Lefkowitz EJ, Lu J, Machi D, Macken C, Mao C,
1125 Niewiadomska A, Nguyen M, Olsen GJ, Overbeek JC, Parrello B, Parrello V, Porter JS, Pusch
1126 GD, Shukla M, Singh I, Stewart L, Tan G, Thomas C, VanOeffelen M, Vonstein V, Wallace ZS,
1127 Warren AS, Wattam AR, Xia F, Yoo H, Zhang Y, Zmasek CM, Scheuermann RH, Stevens RL.
1128 2023. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource
1129 combining PATRIC, IRD and ViPR. Nucleic Acids Res 51:D678–D689.
1130 108. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T,
1131 Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ,
1132 Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren
1133 A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics
1134 Database and Analysis Resource Center. Nucleic Acids Res 45:D535–D542.
1135 109. Federhen S. 2012. The NCBI Taxonomy database. Nucleic Acids Res 40:D136-143.
1136 110. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped
1137 BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids
1138 Res 25:3389–3402.
1139 111. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology
1140 detection and structure prediction. Nucleic Acids Res 33:W244-248.
1141 112. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
1142 BLAST+: architecture and applications. BMC Bioinformatics 10:421.
1143 113. Nielsen H. 2017. Predicting Secretory Proteins with SignalP. Methods Mol Biol Clifton NJ
1144 1611:59–73.
1145 114. Käll L, Krogh A, Sonnhammer ELL. 2004. A combined transmembrane topology and signal
1146 peptide prediction method. J Mol Biol 338:1027–1036.
1147 115. Käll L, Krogh A, Sonnhammer ELL. 2007. Advantages of combined transmembrane topology
1148 and signal peptide prediction--the Phobius web server. Nucleic Acids Res 35:W429-432.

116. Cole C, Barber JD, Barton GJ. 2008. The Jpred 3 secondary structure prediction server. Nucleic Acids Res 36:W197-201.
117. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432.
118. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR. 2015. HMMER web server: 2015 update. Nucleic Acids Res 43:W30-38.
119. Lassmann T. 2019. Kalign 3: multiple sequence alignment of large data sets. Bioinforma Oxf Engl https://doi.org/10.1093/bioinformatics/btz795.
120. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.
121. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PloS One 5:e9490.
122. Rambaut A. 2018. FigTree v1.4.4. GitHub. Repository. https://github.com/rambaut/figtree/releases. Retrieved 13 July 2023.
123. Flouri T, Izquierdo-Carrasco F, Darriba D, Aberer AJ, Nguyen L-T, Minh BQ, Von Haeseler A, Stamatakis A. 2015. The Phylogenetic Likelihood Library. Syst Biol 64:356–362.
124. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Mol Biol Evol 37:291–294.
125. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455.
126. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. 2019. shiny: Web Application Framework for R. https://CRAN.R-project.org/package=shiny.
127. R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
128. RStudio Team. 2020. RStudio: Integrated Development Environment for R. RStudio, PBC., Boston, MA. http://www.rstudio.com/.
129. Hao ZhuThomas Travison DM Timothy Tsai, Will Beasley, Yihui Xie, GuangChuang Yu, Stéphane Laurent, Rob Shepherd, Yoni Sidi, Brian Salzer, George Gui, Yeliang Fan. 2020. Construct Complex Table with “kable” and Pipe Syntax. http://haozhu233.github.io/kableExtra/, https://github.com/haozhu233/kableExtra.
130. Mike Bostock CY Kerry Rodden, Kevin Warne, Kent Russell, Florian Breitwieser. 2020. Sunburst “Htmlwidget.” https://github.com/timelyportfolio/sunburstR.
131. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. 2014. UpSet: Visualization of Intersecting Sets. IEEE Trans Vis Comput Graph 20:1983–1992.
132. Enrico Bonatesta UB Christoph Horejs-Kainrath. 2020. Multiple Sequence Alignment. http://www.bioinf.jku.at/software/msa/.
133. Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinforma Oxf Engl 35:526–528.
134. Fellows I. 2018. Word Clouds. http://blog.fellstat.com/?cat=11.
135. Jake Conway NG. 2019. A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets. http://github.com/hms-dbmi/UpSetR.
136. Müller K. 2017. here: A Simpler Way to Find Your Files. https://CRAN.R-project.org/package=here.

137. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the Tidyverse. J Open Source Softw 4:1686.
138. Martin Morgan MR. 2019. Access the Bioconductor Project Package Repository. https://github.com/Bioconductor/BiocManager.
139. Perry M. 2016. Flexible Heatmaps for Functional Genomics and Sequence Features. https://git.bioconductor.org/packages/heatmaps.
140. Xie Y, Allaire JJ, Grolemund G. 2018. R Markdown: The Definitive Guide. Chapman and Hall/CRC, Boca Raton, Florida. https://bookdown.org/yihui/rmarkdown.
141. Allaire JJ, Xie, Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R. 2020. rmarkdown: Dynamic Documents for R. https://github.com/rstudio/rmarkdown.