



**HAL**  
open science

# Taxonomic distribution of SbmA/BacA and BacA-like antimicrobial peptide transporters suggests independent recruitment and convergent evolution in host-microbe interactions

Nicholas T Smith, Amira Boukherissa, Kiera Antaya, Graeme W Howe, Ricardo C Rodríguez de la Vega, Jacqui A Shykoff, Benoît Alunni, George C Dicenzo

## ► To cite this version:

Nicholas T Smith, Amira Boukherissa, Kiera Antaya, Graeme W Howe, Ricardo C Rodríguez de la Vega, et al.. Taxonomic distribution of SbmA/BacA and BacA-like antimicrobial peptide transporters suggests independent recruitment and convergent evolution in host-microbe interactions. 2024. hal-04803701

**HAL Id: hal-04803701**

**<https://hal.science/hal-04803701v1>**

Preprint submitted on 25 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Title:** Taxonomic distribution of SbmA/BacA and BacA-like antimicrobial peptide transporters  
2 suggests independent recruitment and convergent evolution in host-microbe interactions

3

4 **Authors:** Nicholas T. Smith<sup>1,2+</sup>, Amira Boukherissa<sup>3,4+</sup>, Kiera Antaya<sup>1</sup>, Graeme W. Howe<sup>2</sup>, Ricardo C  
5 Rodríguez de la Vega<sup>4</sup>, Jacqui A. Shykoff<sup>4</sup>, Benoît Alunni<sup>3,5\*</sup>, George C. diCenzo<sup>1\*</sup>

6 + N.T.S. and A.B. contributed equally to this work.

7

8 **Affiliations:** <sup>1</sup> Department of Biology, Queen's University, Kingston, ON, K7L 3N6, Canada

9 <sup>2</sup> Department of Chemistry, Queen's University, Kingston, ON, K7L 3N6, Canada

10 <sup>3</sup> Institute for Integrative Biology of the Cell, CNRS, CEA, Université Paris-Saclay, 91198, Gif-sur-Yvette,  
11 France

12 <sup>4</sup> Écologie Systématique et Évolution, CNRS, Université Paris-Saclay, AgroParisTech, 91198, Gif-sur-  
13 Yvette, France

14 <sup>5</sup> Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin (IJPB), 78000, Versailles,  
15 France

16

17 **\* Corresponding authors:** Benoît Alunni ([benoit.alunni@inrae.fr](mailto:benoit.alunni@inrae.fr)) and George C. diCenzo

18 ([george.dicenzo@queensu.ca](mailto:george.dicenzo@queensu.ca))

19

## ABSTRACT

20 Small, antimicrobial peptides are often produced by eukaryotes to control bacterial populations in  
21 both pathogenic and mutualistic symbioses. These include proline-rich mammalian immune peptides  
22 and cysteine-rich peptides produced by legume plants in symbiosis with rhizobia. The fitness of the  
23 bacterial partner is dependent upon their ability to persist in the presence of these antimicrobial  
24 peptides. In the case of *Escherichia coli* and *Mycobacterium tuberculosis* pathogens and nitrogen-fixing  
25 legume symbionts (rhizobia), the ability to survive exposure to these peptides depends on peptide  
26 transporters called SbmA (also known as BacA) or BclA (for BacA-like). However, how broadly these  
27 transporters are distributed amongst bacteria, and their evolutionary history, is poorly understood.  
28 Here, we used hidden Markov models, phylogenetic analysis, and sequence similarity networks to  
29 examine the distribution of SbmA/BacA and BclA proteins across a representative set of 1,255 species  
30 from across the domain *Bacteria*. We identified a total of 71 and 177 SbmA/BacA and BclA proteins,  
31 respectively. Phylogenetic and sequence similarity analyses suggest that these protein families likely  
32 did not evolve from a common ancestor and that their functional similarity is instead a result of  
33 convergent evolution. *In vitro* sensitivity assays using the legume peptide NCR247 and several of the  
34 newly-identified BclA proteins confirmed that transport of antimicrobial peptides is a common feature  
35 of this protein family. Analysis of the taxonomic distribution of these proteins showed that SbmA/BacA  
36 orthologs were encoded only by species in the phylum *Pseudomonadota* and that they were primarily  
37 identified in just two orders: *Hyphomicrobiales* (class *Alphaproteobacteria*) and *Enterobacterales* (class  
38 *Gammaproteobacteria*). BclA orthologs were somewhat more broadly distributed and were found in  
39 clusters across four phyla. These included several orders of the phyla *Pseudomonadota* and  
40 *Cyanobacteriota*, as well as the order *Mycobacteriales* (phylum *Actinomycetota*) and the class  
41 *Negativicutes* (phylum *Bacillota*). Notably, many of the clades enriched for species encoding BacA or  
42 BclA orthologs also include many species known to interact with eukaryotic hosts in mutualistic or  
43 pathogenic interactions. Collectively, these observations suggest that SbmA/BacA and BclA proteins  
44 have been repeatedly co-opted to facilitate both mutualistic and pathogenic associations with  
45 eukaryotic hosts by allowing bacteria to cope with host-encoded antimicrobial peptides.

46

47 **Keywords:** SbmA, antimicrobial peptides, host-microbe interaction, peptide transport, convergent  
48 evolution, rhizobium-legume symbioses, pathogenesis

49

## INTRODUCTION

50 Rhizobia are a polyphyletic group of bacteria from the classes *Alphaproteobacteria* and  
51 *Betaproteobacteria* that can enter endosymbiotic relationships with leguminous plants. During the  
52 symbiosis, rhizobia reside within cells of a specialized structure called a root nodule, where the  
53 bacteria convert atmospheric N<sub>2</sub> gas into ammonium through a process known as symbiotic nitrogen  
54 fixation (SNF). The fixed nitrogen is provided to the plant host in exchange for photosynthetically fixed  
55 carbon, allowing plant growth in otherwise nitrogen-limited conditions (1). As a result, SNF is  
56 frequently leveraged in agriculture in place of nitrogen fertilizers, which are both expensive and a  
57 major source of agricultural greenhouse gas emissions (2, 3).

58 Within nodules of legumes of the inverted-repeat lacking clade (IRLC) and Dalbergioid clade,  
59 rhizobia undergo a process called terminal bacteroid differentiation (TBD) prior to fixing nitrogen (4,  
60 5) This involves bacterial cell enlargement, genome endoreduplication, increased membrane  
61 permeability, and major changes in gene expression and cellular metabolism (6–8). TBD is essential  
62 for SNF in these legumes and is driven by a family of small, plant-encoded proteins known as nodule-  
63 specific cysteine-rich (NCR) peptides, which target the intracellular symbiotic bacteria to promote TBD  
64 (9, 10). IRLC and Dalbergioid legumes encode between <10 and >700 NCR peptides that contain four,  
65 six, or eight conserved cysteine residues (11, 12) but otherwise share little sequence conservation and  
66 have isoelectric points (pI) that vary from 3 (anionic) to 11 (cationic) (11, 13). Cationic NCR peptides  
67 with pI values above nine display antimicrobial activity, likely due to interaction with the negatively-  
68 charged surface of bacterial plasma membranes resulting in membrane depolarization (14, 15).  
69 During symbioses with rhizobia, NCR peptides are thought to additionally interact with rhizobial  
70 cytosolic proteins, such as cell cycle regulators and cell division proteins, to promote TBD (15, 16).

71 NCR peptide-triggered TBD involves bacteria-encoded ABC transporters known as BacA (a  
72 homolog of *Escherichia coli* SbmA) and BclA (BacA-like); all rhizobial symbionts of IRLC and Dalbergioid  
73 legumes encode either a BacA or BclA protein, but to our knowledge, not both. BacA and BclA are  
74 inner membrane peptide transporters, differing primarily by the presence of an ATPase domain in  
75 BclA that is absent in BacA (17). ATP hydrolysis by the ATPase domain is essential for the transport  
76 activity of BclA, whereas BacA-mediated transport is driven by the proton-motive force (18, 19).  
77 Despite this difference, both BacA and BclA can import NCR peptides, and their loss renders rhizobia  
78 hypersensitive to NCR peptide exposure *in vitro* (17). Loss-of-function mutation of *bacA* or *bclA* results  
79 in several other free-living phenotypes, including increased resistance to the antibiotics bleomycin  
80 and gentamicin and, in some species, increased sensitivity to detergents and altered

81 lipopolysaccharide modifications (20–26). In addition, rhizobia carrying loss-of-function *bacA* or *bclA*  
82 mutations rapidly die upon release into the nodules of IRLC or Dalbergioid legumes in an NCR peptide-  
83 dependent fashion and thus fail to differentiate (17, 27). On the other hand, *bacA* and *bclA* are not  
84 required for symbiosis with legumes that produce no NCR peptides (21). It has been hypothesized that  
85 the requirement of BacA or BclA for SNF in legume plants that induce TBD may be two-fold: (i) to move  
86 NCR peptides away from the cell membrane, thereby protecting the bacteria from their antimicrobial  
87 activities, and (ii) to transport the NCR peptides to their intracellular targets to promote TBD (17, 28,  
88 29). While BacA and BclA are the primary transporters of NCR peptides in all rhizobia for which this  
89 trait has been studied, recent results suggest that another broad-specificity ABC transporter, the  
90 *Sinorhizobium meliloti* YejABEF protein, may also contribute to the import of NCR peptides and  
91 bacterial survival when grown in the presence of NCR peptides (30).

92 Orthologs of BacA and BclA have also been found in multiple pathogenic bacteria, including *E.*  
93 *coli* (where it is known as SbmA) (31), *Brucella abortus* (32), and *Mycobacterium tuberculosis* (33). Similar  
94 to how BacA or BclA is required for rhizobial survival in nodules of legume plants inducing TBD, BacA  
95 and BclA orthologs are required by *B. abortus* and *M. tuberculosis* for chronic infection of their  
96 eukaryotic hosts (32–34), likely by providing resistance to host-encoded antimicrobial peptides as  
97 these transporters are required for peptide transport and resistance *in vitro* (18, 19). Despite the  
98 conserved role of BacA and BclA as peptide transporters essential for beneficial and pathogenic  
99 interactions, orthologs vary in their ability to functionally replace each other. Whereas the *bacA* genes  
100 of *B. abortus* and *S. meliloti* complement the symbiotic defect of a *S. meliloti bacA* null mutant (20, 35),  
101 little to no complementation of the symbiotic defects is observed when the *S. fredii*, *R. leguminosarum*,  
102 or *Mesorhizobium loti bacA* are expressed in an *S. meliloti bacA* null mutant (23, 28). Likewise, the *bclA*  
103 genes of *M. tuberculosis* and *Bradyrhizobium* spp. cannot restore nitrogen fixation when expressed in  
104 a *S. meliloti bacA* null mutant (33, 36, 37). However, despite the lack of complementation of symbiotic  
105 phenotypes, most *bacA* and *bclA* orthologs still complement the gentamicin and/or bleomycin  
106 phenotypes of *S. meliloti bacA* null mutants (20, 23, 28, 33, 35–37). These results suggest that BacA and  
107 BclA orthologs display slight variations in their peptide substrate range or rate of transport (13).

108 The observation that BacA and BclA orthologs are found in diverse bacterial lineages suggests  
109 that these proteins may be widespread housekeeping proteins subsequently co-opted for host-  
110 bacterial interactions (38). However, no systematic study of the distribution of BacA and BclA orthologs  
111 on the bacterial tree exists. In addition, the evolutionary relationship between the BacA and BclA  
112 families remains to be elucidated. Here, we report the distribution of BacA and BclA orthologs in 1,255

113 bacterial species from across the bacterial domain. We found BacA orthologs exclusively within the  
114 phylum *Pseudomonadales* (syn. *Proteobacteria*), while BclA orthologs were predominately limited to  
115 the phyla *Pseudomonadales*, *Cyanobacteriota* (syn. *Cyanobacteria*), *Actinomycetota* (syn. *Actinomycetes*),  
116 and *Bacillota* (syn. *Firmicutes*). Expression of a subset of the newly identified BclA proteins in *S. meliloti*  
117  $\Delta bacA$  mutants confirmed that transport of antimicrobial peptides is a common property of the BclA  
118 protein family. The taxonomic distribution of SbmA/BacA and BclA, together with phylogenetic  
119 analysis of these proteins, leads us to suggest that the functional similarities between SbmA/BacA and  
120 BclA are a result of convergent evolution, and that these protein families have been repeatedly co-  
121 opted to help microbes cope with antimicrobial peptide exposure during host-microbe interactions.

122

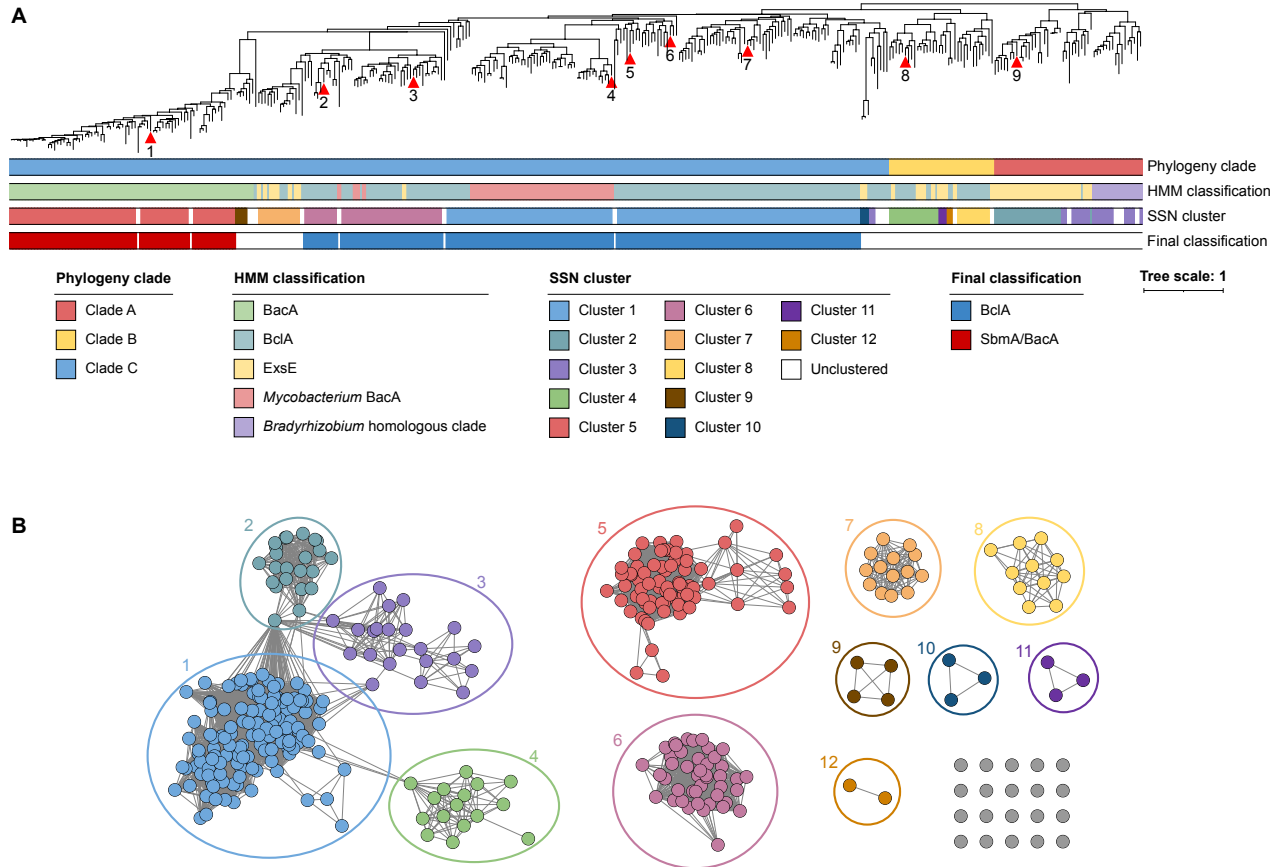
123

## RESULTS

### 124 Identification and classification of SbmA/BacA and BclA orthologs across the bacterial domain

125 To study the evolution and distribution of SbmA/BacA and BclA proteins, we searched the proteomes  
126 of 1,255 bacterial species, each belonging to a distinct genus, for proteins showing similarity to the  
127 SbmA/BacA-like family of PFAM (PF05992) (see Materials and Methods). This process led to the  
128 identification of 366 putative SbmA/BacA-like family proteins from 258 species. We further classified  
129 each of these 366 proteins into one of five protein classes according to Guefrachi and colleagues (36):  
130 SbmA/BacA, BclA, *Mycobacterium* BacA (a BclA-like family of proteins first identified in *M. tuberculosis*),  
131 ExsE (a related protein family involved in long-chain fatty acid transport), and the so-called  
132 *Bradyrhizobium* homologous clade (a related protein family with an unknown function). Initially, this  
133 classification was based on the use of hidden Markov models (HMMs), which was subsequently  
134 refined based on phylogenetic reconstruction and a sequence similarity network (SSN) as described  
135 below.

136 Using HMMs for these five protein classes, the 366 SbmA/BacA-like family proteins were initially  
137 classified into 79 SbmA/BacA proteins, 169 BclA proteins, 50 *Mycobacterium* BacA proteins, 52 ExsE  
138 proteins, and 16 *Bradyrhizobium* homologous clade proteins (**Figure 1A**). A maximum-likelihood  
139 phylogenetic analysis led to the identification of three primary monophyletic groups (**Figure 1A**). Clade  
140 A comprised 48 proteins and included most ExsE and all *Bradyrhizobium* homologous clade proteins,  
141 which we treated as the outgroup. Clade B included 34 proteins that were annotated as a mix of BclA  
142 and ExsE based on the HMMs. Clade C was the largest clade consisting of 284 proteins, and included  
143 most of the putative BclA, SbmA/BacA, and *Mycobacterium* BacA proteins.



144  
145

**Figure 1. Sequence and phylogenetic analysis of SbmA/BacA-like proteins.** (A) An unrooted maximum likelihood phylogeny of 366 SbmA/BacA-like proteins is shown. The scale bar represents the average number of amino acid substitutions per site. Red triangles indicate proteins whose corresponding genes were codon optimized and synthesized: 1 - *Polymorphum gilvum* BclA; 2 - *Synechococcus elongatus* BclA; 3 - *Cyanobacterium aponinum* BclA; 4 - *Basilea psittacipulmonis* BclA; 5 - *Succinivibrio dextrinosolvens* BclA; 6 - *Methylomusa anaerophila* BclA; 7 - *Polaromonas naphthalenivorans* BclA; 8 - *Eikenella exigua* BclA-like; 9 - *Phyllobacterium zundukense* ExsE. The bars beneath the phylogeny summarize the clustering and annotation of these proteins. The top bar indicates the phylogenetic clade to which each protein belongs. The second bar indicates the preliminary hidden Markov model (HMM) classification of each protein. The third bar indicates the cluster in the sequence similarity network that each protein belongs to. The bottom bar indicates which proteins were ultimately classified as SbmA/BacA (red) or BclA (blue). An interactive version of this phylogeny, with node support values, is provided through iTol (<https://itol.embl.de/shared/11AjjFrHYGLI9>) while a Newick-formatted version of the phylogeny can be downloaded from GitHub ([https://github.com/amira-boukh/SbmA\\_BacA\\_phylogenetic\\_distribution](https://github.com/amira-boukh/SbmA_BacA_phylogenetic_distribution)). (B) A sequence similarity network, calculated using EFl-EST, of 366 SbmA-BacA-like proteins is shown. Each node (the circles) represents one protein, while edges (the lines) represent sequence similarity between pairs of proteins above the threshold, with longer lines indicating lower similarity. Nodes are colour coded based on cluster.



164 Most of the putative BclA proteins from Clade C also form a single cluster in the SSN (Cluster 1;  
165 **Figure 1B**). We therefore conclude that the 133 proteins of Cluster 1 in the SSN represent true BclA  
166 orthologs. Notably, Cluster 1 of the SSN also includes 46 proteins annotated as *Mycobacterium* BacA,  
167 which also fall within Clade C in the phylogeny (**Figure 1**). This suggests that the *Mycobacterium* BacA  
168 proteins are not a distinct family from the BclA proteins, and that *Mycobacterium* BacA proteins should  
169 instead be referred to as BclA. On the other hand, a Clade C subclade of nine proteins with long branch  
170 lengths in the phylogeny is excluded from Cluster 1 of the SSN; instead, two of these proteins are  
171 found as part of Cluster 3 that predominantly consists of the *Bradyrhizobium* homologous clade  
172 proteins, three are found as a three-protein cluster (Cluster 10), and five are singletons. In addition,  
173 four of these nine proteins are from strains encoding a BclA protein belonging to Cluster 1. Taken  
174 together, we conclude that these nine proteins are not true BclA orthologs. Another subclade of Clade  
175 C consisting of 58 proteins is not part of Cluster 1 in the SSN but rather is largely found in two clusters  
176 (Clusters 6 and 7) of 44 and 14 proteins, respectively (**Figure 1**). Cluster 6 consists primarily of proteins  
177 from cyanobacteria, and 43 of the 44 proteins were classified as BclA or *Mycobacterium* BacA by the  
178 HMMs. In addition, the functional data described below suggests proteins of this cluster are  
179 functionally similar to known BclA proteins. We therefore conclude that proteins of Cluster 6 represent  
180 BclA orthologs. In contrast, eight of the 14 proteins of Cluster 7 were annotated as ExsE by the HMMs.  
181 The distinct clustering of Cluster 7 from Cluster 6, together with the HMM annotations, lead us to  
182 suggest that the proteins of Cluster 5 are unlikely to represent true BclA orthologs.

183 Consistent with the phylogenetic analysis, proteins of Clade B do not cluster with proteins of  
184 Clade C in the SSN (**Figure 1B**). Rather, the Clade B proteins are split across four clusters and two  
185 singletons. Nearly 1/3<sup>rd</sup> (10 of 34) proteins of Clade B were annotated as ExsE by the initial HMM  
186 strategy, and many of the proteins of Clade B are from bacterial strains that also encode a putative  
187 SbmA/BacA or BclA of Clade C. Collectively, we interpret these results to indicate that Clade B proteins  
188 are not part of the BclA protein family, and that they instead represent a related but distinct protein  
189 family. This conclusion is also supported by the functional data presented below.

190 Lastly, all putative SbmA/BacA proteins formed a monophyletic group in the phylogeny (**Figure**  
191 **1A**) and a monophyletic group of 71 of the 79 proteins form a single cluster (Cluster 5) in the SSN  
192 (**Figure 1B**). These results suggest that the 71 proteins of Cluster 5 and annotated as SbmA/BacA by  
193 the HMM strategy are likely true SbmA/BacA orthologs, and that all SbmA/BacA proteins evolved from  
194 a common ancestor. Although the SbmA/BacA proteins fell within Clade C in the phylogeny, the  
195 SbmA/BacA clade is connected to the rest of the tree via an unusually long branch, consistent with the



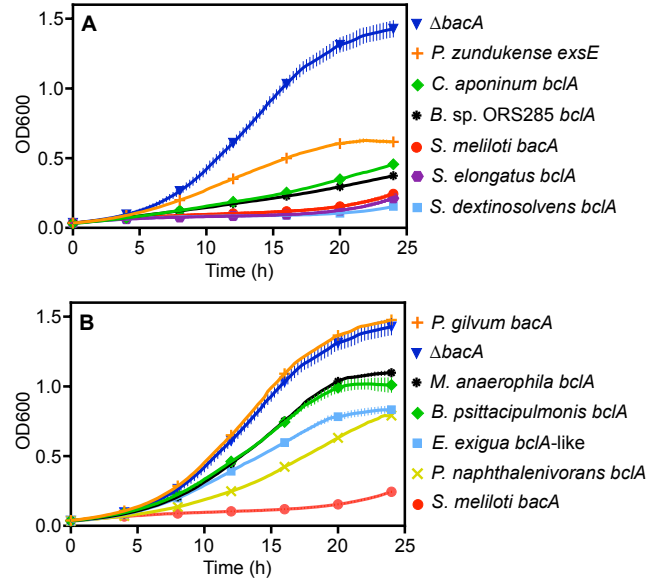
196 distinct clustering of SbmA/BacA proteins in the SSN. The distinct clustering in the SSN, the long  
197 branch length, and the functional differences in transport (ATP-driven vs proton-driven) lead us to  
198 suggest that the SbmA/BacA and BclA protein families evolved independently and that their functional  
199 similarity is a result of convergent evolution.

200 In considering the different sources of information described above, we ultimately chose to  
201 select a final set of SbmA/BacA and BclA proteins based primarily on the SSN, resulting in the  
202 identification of 177 high-confidence BclA proteins (including the *Mycobacterium* BacA proteins) and  
203 71 high-confidence BacA proteins (**Figure 1A**).

204

### 205 ***In vitro* functional analysis of diverse SbmA/BacA and BclA orthologs**

206 To validate that the BclA and SbmA/BacA proteins identified through the *in silico* approach are  
207 functionally similar to known BclA and SbmA/BacA proteins, genes encoding nine of the identified  
208 proteins were synthesized. The proteins encoded by these genes included: one BacA protein, six BclA  
209 proteins including one previously classified as *Mycobacterium* BacA, one protein from Clade B  
210 (henceforth referred to as BacA-like), and one ExsE protein for comparison. The nine genes were then  
211 cloned into an expression vector and introduced into *S. meliloti*  $\Delta bacA$  and *S. meliloti*  $\Delta bacA$   $\Omega yejA$   
212 mutants to test for complementation. Although the genes were codon-optimized for expression in *S.*  
213 *meliloti*, we cannot exclude the possibility that some proteins were not properly expressed or were  
214 not stably inserted into the *S. meliloti* inner membrane. Therefore, lack of complementation may  
215 reflect improper expression/localization of a protein rather than a lack of orthology. All strains showed  
216 similar growth in media lacking antimicrobial agents (**Figure S1**), indicating that differences in media  
217 supplemented with gentamicin (Gm) or NCR peptides reflect altered resistance phenotypes rather  
218 than general growth differences. In addition, we observed that the resistance phenotypes of the *S.*  
219 *meliloti*  $\Delta bacA$  mutant complemented with the *S. meliloti* *bacA* gene *in trans* differed somewhat from  
220 wildtype *S. meliloti* (**Figure S2**), likely due to elevated expression of *bacA* in the complemented strain.  
221 Thus, for all *in vitro* phenotypic experiments, strains were compared to the *S. meliloti*  $\Delta bacA$  mutant  
222 complemented with the *S. meliloti* *bacA* gene *in trans* rather than the wild type.



223  
224

225 **Figure 2. Gentamicin sensitivity assays.** The growth of various *S. meliloti* strains, as measured by  
226 OD600, in the presence of 20 µg/mL of gentamicin is shown over a 24-hour period. Each point  
227 represents the mean of triplicate wells, with error bars depicting standard deviation. The  $\Delta bacA$  strain  
228 represents the *S. meliloti*  $\Delta bacA$  mutant carrying an empty vector, while all other strains are named  
229 according to the species of origin of the gene expressed *in trans* in the *S. meliloti*  $\Delta bacA$  background.  
230 The experiment was replicated three independent times, and data from a representative experiment  
231 is shown. (A) Data is shown for genes exhibiting moderate to high level of complementation of the *S.*  
232 *meliloti*  $\Delta bacA$  gentamicin resistance phenotype. (B) Data is shown for genes exhibiting low to  
233 moderate levels of complementation of the *S. meliloti*  $\Delta bacA$  gentamicin resistance phenotype.

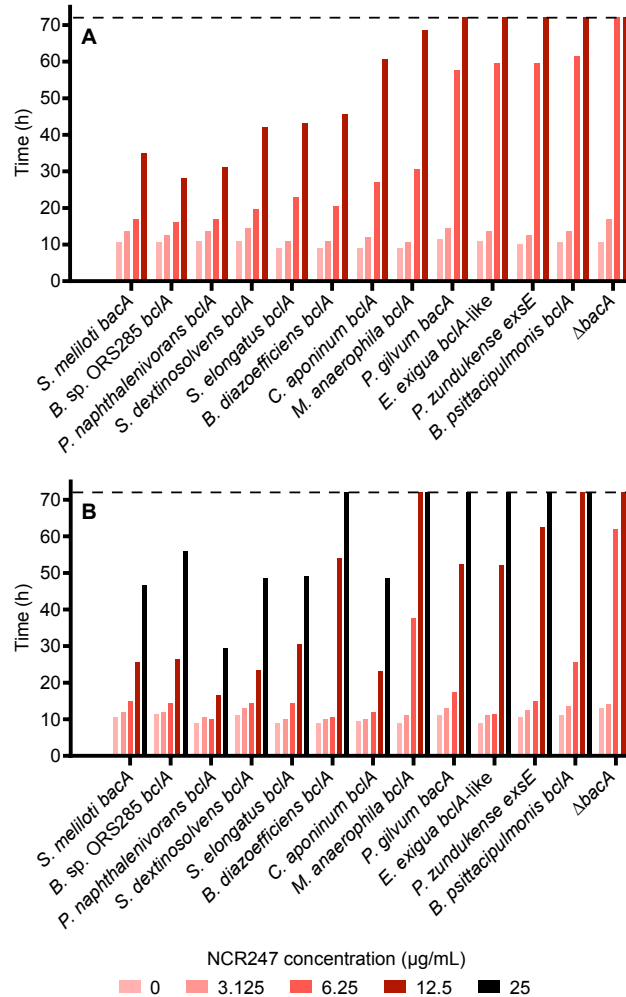
234

235

236 We first tested whether the nine genes could complement the Gm resistance phenotype of the  
237 *S. meliloti*  $\Delta bacA$  mutant. As expected, the  $\Delta bacA$  mutant was resistant to Gm, and reintroduction of  
238 the *S. meliloti* *bacA* gene *in trans* resulted in sensitivity to Gm (Figure 2A). Unexpectedly, introduction  
239 of the *Phyllobacterium zundukense* *exsE* gene resulted in intermediate complementation of the Gm  
240 resistance phenotype (Figure 2A), suggesting that transport of Gm is a broadly conserved function of  
241 the SbmA/BacA and related proteins, and is not specific to BclA or SbmA/BacA proteins. As a result,  
242 the impact of the nine genes on Gm resistance cannot be used to support the annotation of a protein  
243 specifically as BclA or SbmA/BacA; however, it is still a useful metric to test whether a SbmA/BacA-like  
244 protein is expressed and functional. Of the six *bclA* genes identified by our screen, three (from  
245 *Cyanobacterium aponinum*, *Synechococcus elongatus*, and *Succinivibrio dextrinosolvens*) complemented  
246 the Gm resistance phenotype at least as well as the known *bclA* gene of *Bradyrhizobium* sp. ORS285  
247 (Figure 2A), confirming they are expressed and functional in *S. meliloti*. The other three *bclA* genes all

248 displayed partial complementation to varying degrees (**Figure 2B**), suggesting they are expressed and  
249 functional but either have reduced ability to transport Gm or their expression or stability is sub-  
250 optimal. Likewise, the one BclA-like gene (from *Eikenella exigua*) displayed partial complementation of  
251 the Gm resistance phenotype (**Figure 2B**). On the other hand, introduction of the one *bacA* gene that  
252 we tested (from *Polymorphum gilvum*) completely failed to complement the Gm resistance phenotype  
253 of the *S. meliloti*  $\Delta bacA$  mutant (**Figure 2B**), which we hypothesize is due to improper expression or  
254 stability of the protein rather than functional divergence.

255 We next indirectly examined whether the nine proteins could transport eukaryotic antimicrobial  
256 peptides by measuring the impact of the proteins on the sensitivity of *S. meliloti* to the legume-  
257 encoded NCR peptide NCR247 (**Figure 3**); proteins transporting NCR247 are expected to show  
258 reduced sensitivity to this peptide. As expected, the *S. meliloti*  $\Delta bacA$  single mutant and the  $\Delta bacA$   $\Omega yejA$   
259 double mutant were hypersensitive to NCR247 exposure, while introduction of the known *S. meliloti*  
260 *bacA* or *Bradyrhizobium* sp. ORS285 *bclA* genes *in trans* resulted in reduced sensitivity to NCR247  
261 (**Figure 3**). Introduction of the *P. zundukense* *exsE* gene into the two mutants resulted in little to no  
262 complementation of the NCR247 hypersensitivity phenotypes (**Figure 3**), consistent with the transport  
263 of NCR peptides being specific to the SbmA/BacA and BclA family proteins and not a general property  
264 of these and related proteins. All three of the *bclA* genes showing strong complementation of the Gm  
265 resistance phenotype (two of which are from cyanobacteria) also showed good complementation of  
266 the NCR247 hypersensitivity phenotype (**Figure 3**), confirming the proteins encoded by these three  
267 genes are functionally similar to known BclA proteins. In addition, the *bclA* gene from *P.*  
268 *naphthalenivorans* strongly complemented the NCR247 hypersensitivity phenotypes of both strains  
269 despite only moderate complementation of the Gm resistance phenotype. Of the remaining two *bclA*  
270 genes, one (from *Methylophila anaerophila*) displayed weak complementation of the NCR247  
271 hypersensitivity (**Figure 3**) and varied in its level of complementation across trials (not shown), while  
272 one (from *Basilea psittacipulmonis*) failed to complement (**Figure 3**). Overall, the data for the six BclA  
273 proteins support that most BclA proteins are capable of transporting NCR peptides. On the other  
274 hand, the NCR247 sensitivity phenotypes of the strains expressing the BclA-like protein from *E. exigua*  
275 resembled the phenotypes of the strain expressing *P. zundukense* *exsE* (**Figure 3**), consistent with BclA-  
276 like proteins of Clade B (**Figure 1**) representing a different class of proteins from BclA. In accordance  
277 with the Gm resistance data, the *bacA* gene from *P. gilvum* largely failed to complement the NCR247  
278 hypersensitivity phenotypes (**Figure 3**), potentially reflecting improper expression or stability of the  
279 encoded protein.



280  
281

282 **Figure 3. NCR247 sensitivity assays.** The growth of various *S. meliloti* strains, as measured by OD600,  
283 in the presence of the antimicrobial peptide NCR247 is shown. Strains were grown in various  
284 concentrations of NCR247 as indicated by the shade of red or black. Bars represent the time required  
285 for the culture to reach an OD600 of 0.25. Values of 72 hours (indicated by the dashed line) indicate  
286 that the strain failed to reach an OD600 of 0.25 within the 72-hour growth period. The  $\Delta bacA$  label  
287 represents the *S. meliloti* (A)  $\Delta bacA$  or (B)  $\Delta bacA$  *QyejA* mutant carrying an empty vector, while all other  
288 strains are named according to the species of origin of the gene expressed *in trans* in the *S. meliloti* (A)  
289  $\Delta bacA$  or (B)  $\Delta bacA$  *QyejA* background. (A) Data is shown for the *S. meliloti*  $\Delta bacA$  mutant and  
290 derivatives. (B) Data is shown for the *S. meliloti*  $\Delta bacA$  *QyejA* mutant and derivatives.

291

292

### 293 Analysis of the ability of BacA and BclA to support legume symbiosis

294 We additionally tested whether the nine proteins could complement the nitrogen-fixation defect of a  
295 *S. meliloti*  $\Delta bacA$  mutant during symbiosis with *Medicago sativa* (alfalfa) or *Melilotus officinalis* (yellow-  
296 blossom sweet clover). As expected, the *S. meliloti*  $\Delta bacA$  mutant formed small white nodules on both  
297 plants and failed to fix nitrogen, while re-introduction of the *S. meliloti* *bacA* gene *in trans*

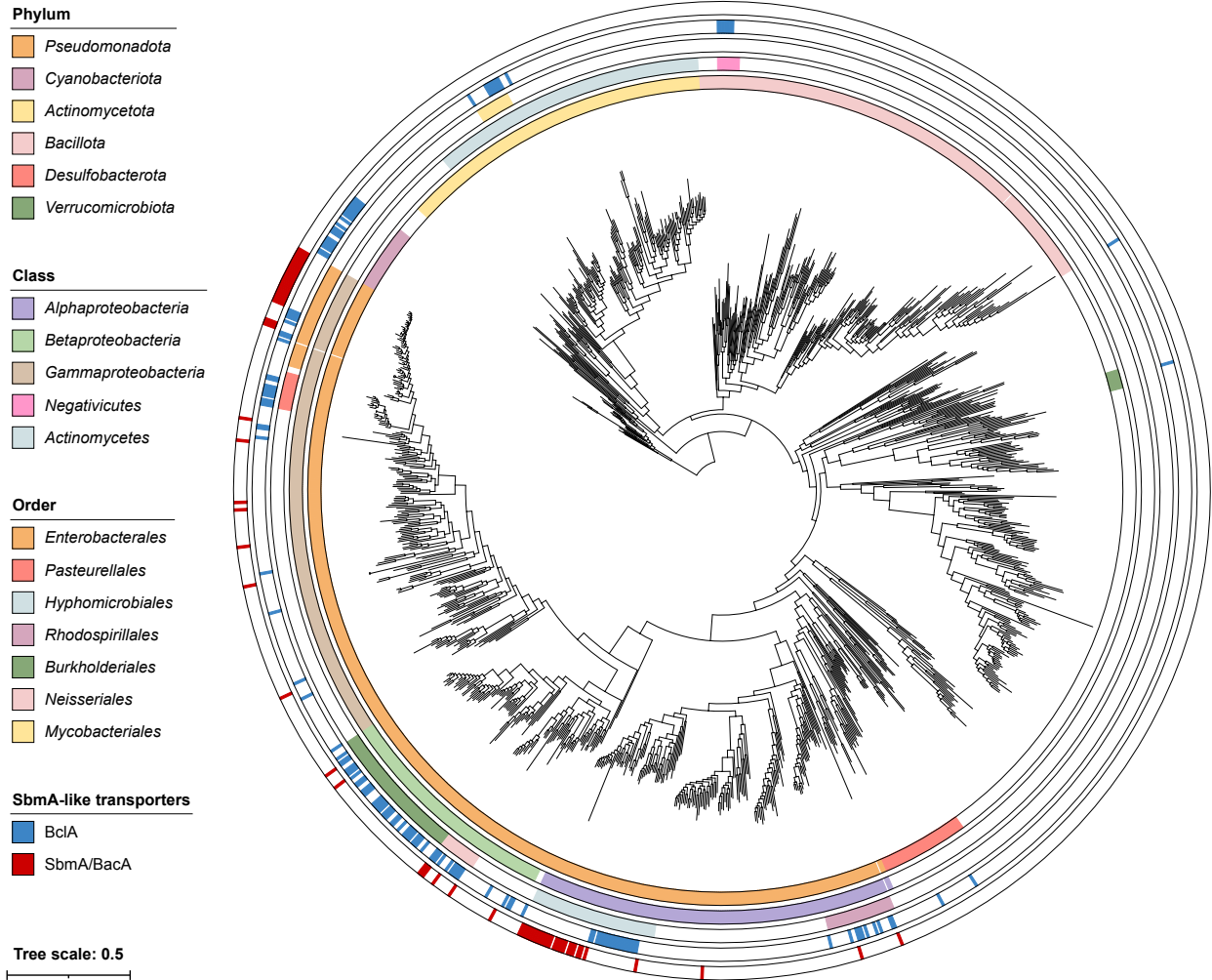
298 complemented the nitrogen-fixation phenotype (**Table S1**). All nine of the synthesized genes failed to  
299 complement the nitrogen-fixation phenotype (**Table S1**). As the same lack of complementation was  
300 observed for the known *bclA* gene of *Bradyrhizobium* sp. ORS285 (**Table S1**), these results suggest that  
301 most, if not all, BclA proteins are unable to support an effective symbiosis between *S. meliloti* and its  
302 host plants.

303

#### 304 **Taxonomic distribution of SbmA/BacA and BclA orthologs across the domain *Bacteria***

305 We next examined the taxonomic distribution of the 177 BclA and 71 SbmA/BacA proteins identified  
306 as described earlier. Remarkably, 100% and 78% of the identified SbmA/BacA and BclA proteins,  
307 respectively, are encoded by species of the phylum *Pseudomonadales* (syn. *Proteobacteria*) (**Figure 4**).  
308 As expected, most species encoding SbmA/BacA or BclA proteins encode only one or the other; only  
309 six of the 208 species encoding SbmA/BacA and/or BclA encode both, and in all six cases, both genes  
310 are carried by the chromosome.

311       Approximately 70% of the SbmA/BacA proteins are encoded by just two monophyletic groups  
312 of organisms, suggesting that SbmA/BacA was acquired at the base of each clade and then vertically  
313 transmitted. These two clades are a 24 species clade in the order *Enterobacterales* (all of which encode  
314 BacA) and a 29 species clade in the order *Hyphomicrobiales* (25 of which encode SbmA/BacA) (**Figure**  
315 **4**). Interestingly, the SbmA/BacA proteins of the order *Enterobacterales* form a monophyletic group in  
316 the SbmA/BclA protein phylogeny (**Figure S3**). On the other hand, the minimal monophyletic clade  
317 encompassing all *Hyphomicrobiales* SbmA/BacA proteins also includes the *Enterobacterales*  
318 SbmA/BacA proteins (**Figure S3**). These results suggest that SbmA/BacA proteins of the order  
319 *Enterobacterales* were acquired through horizontal transfer from the order *Hyphomicrobiales*. The  
320 remaining 22 SbmA/BacA proteins not found within those two clades are distributed across the  
321 phylum *Pseudomonadales* with no other major clustering observed. Overall, these results suggest that  
322 although SbmA/BacA proteins are widespread amongst subclasses of the orders *Enterobacterales* (class  
323 *Gammaproteobacteria*) and *Hyphomicrobiales* (class *Alphaproteobacteria*), the taxonomic distribution of  
324 this protein family is otherwise limited.



325  
326

327 **Figure 4. Taxonomic distribution of SbmA/BacA and BclA proteins in the domain *Bacteria*.** An  
328 unrooted maximum likelihood phylogeny of 1,533 bacteria is shown, inferred from the concatenated  
329 protein alignments of 31 single-copy proteins. The scale bar represents the average number of amino  
330 acid substitutions per site. Three clades of intracellular symbionts/pathogens with long branch lengths  
331 were removed for presentation purposes; none of these taxa encode SbmA/BacA or BclA. The outer  
332 rings represent the following, starting from the inner ring: (i) the phylum that each strain belongs to,  
333 limited to phyla where at least one strain encodes SbmA/BacA or BclA; (ii) the class that each strain  
334 belongs to, limited to classes where at least one strain encodes SbmA/BacA or BclA and that are  
335 mentioned in the text; (iii) the class that each strain belongs to, limited to classes where at least one  
336 strain encodes SbmA/BacA or BclA and that are mentioned in the text; (iv) whether the strain encodes  
337 BclA (blue) or not (white); (v) whether the strain encodes SbmA/BacA (red) or not (white). An interactive  
338 version of this phylogeny, with node support values and without collapsing of any clades, is provided  
339 through iTol (<https://itol.embl.de/shared/11AjjFrHYGLI9>) while a Newick-formatted version of the  
340 phylogeny can be downloaded from GitHub ([https://github.com/amira-  
341 boukh/SbmA\\_BacA\\_phylogenetic\\_distribution](https://github.com/amira-boukh/SbmA_BacA_phylogenetic_distribution)).



342 BclA proteins show a somewhat broader taxonomic distribution than the SbmA/BacA proteins,  
343 although their distribution remains restricted to only a few phyla (**Figure 4**). Like SbmA/BacA, BclA was  
344 common in a subclade of the order *Hyphomicrobiales*, in which 19 of 21 species encoded BclA. Most  
345 of the other *Alphaproteobacteria* species encoding BclA belong to the order *Rhodospirillales*, in which  
346 nine of the 30 species encoded BclA. Within the *Gammaproteobacteria*, the taxon most enriched for  
347 BclA proteins was the order *Pasteurellales*, in which 10 of the 16 species encoded BclA. BclA was also  
348 abundant in the class *Betaproteobacteria*, unlike SbmA/BacA, and was particularly enriched in the  
349 orders *Burkholderiales* (32/60 species) and *Neisseriales* (10/17 species) compared to the orders  
350 *Nitrosomonadales* and *Rhodocyclales* (4/27 species across both orders). In contrast to SbmA/BacA,  
351 which was predicted to be encoded only by species of the phylum *Pseudomonadales*, there were three  
352 main clades of organisms predicted to encode BclA outside of the phylum *Pseudomonadales* (**Figure**  
353 **4**). The largest of these were the phylum *Cyanobacteriota* (syn. *Cyanobacteria*), in which BclA was  
354 broadly distributed and found in 21 of the 31 species (~67%). The other two main groups of organisms  
355 encoding BclA are a subclade of eight species (seven of which encode BclA) of the order  
356 *Mycobacteriales* (phylum *Actinomycetota* [syn. *Actinomycetes*]), and the class *Negativicutes* (phylum  
357 *Bacillota* [syn. *Firmicutes*]) in which seven of the ten species encode BclA orthologs (**Figure 4**).

358  
359

## DISCUSSION

360 We identified 71 SbmA/BacA and 177 BclA orthologs from a search of the proteomes of 1,255 bacterial  
361 species. In total, 208 of the 1,255 species (16.6%) encoded at least one copy of SbmA/BacA and/or  
362 BclA, with only six of the 208 species (2.9%) encoding both SbmA/BacA and BclA. The observation that  
363 SbmA/BacA and BclA proteins were generally not encoded in the same proteome suggests that these  
364 protein families have similar biological roles. We also observed that the so-called “*Mycobacterium*  
365 BacA” proteins clustered with the BclA proteins in both the SSN and the protein phylogeny, leading us  
366 to conclude that the “*Mycobacterium* BacA” proteins are not distinct from BclA; we therefore  
367 reclassified the “*Mycobacterium* BacA” proteins as BclA for downstream analyses.

368

### 369 **Convergent evolution of the SmbA/BacA and BclA protein families**

370 One of the objectives motivating this work was to gain insight into whether the SbmA/BacA and BclA  
371 protein families share common ancestry (e.g., that SbmA/BacA evolved from BclA, or vice versa) or  
372 whether they evolved independently and converged towards a similar function. The taxonomic  
373 distribution of SbmA/BacA and BclA proteins within the order *Hyphomicrobiales* is potentially



374 suggestive of the former scenario. Excluding the deep-branching lineages, the order *Hyphomicrobiales*  
375 can be sub-divided into two sister clades; SbmA/BacA is widely distributed in one of these clades, while  
376 BclA is widely distributed in the other. This could suggest that the SbmA/BacA and BclA proteins of the  
377 order *Hyphomicrobiales* evolved from a common ancestral protein present in the ancestor of these  
378 clades. However, the *Hyphomicrobiales* SbmA/BacA and BclA proteins are polyphyletic in the BacA/BclA  
379 protein phylogeny, which instead suggests that the SbmA/BacA and BclA proteins of the order  
380 *Hyphomicrobiales* were independently acquired. The distinct clustering of the BclA and SbmA/BacA  
381 proteins in the SSN further supports independent evolutionary origins for these proteins, as does the  
382 notably long branch connecting the SbmA/BacA clade to the rest of the phylogeny. Moreover, we  
383 consider the differences in transport mechanisms of SbmA/BacA (proton gradient-driven) and BclA  
384 (ATP-driven) to be more easily explained if these protein families have separate evolutionary histories.  
385 Overall, we interpret the evidence as suggesting that the SbmA/BacA and BclA protein families evolved  
386 independently, and that their functional similarity is a result of convergent molecular evolution.

387 Twenty-eight of the BclA proteins were encoded by 21 cyanobacteria. These 28 proteins formed  
388 a distinct cluster in the SSN together with 15 non-cyanobacterial BclA proteins, raising the possibility  
389 that these proteins also evolved independently from the rest of the BclA proteins. While we cannot  
390 rule out this possibility, we consider the evidence to be insufficient to reach this conclusion at this  
391 time.

392

### 393 **The SbmA/BacA and BclA protein families are associated with eukaryotic host interaction**

394 A second objective of this work was to determine how broadly SbmA/BacA and BclA proteins are  
395 distributed across the domain *Bacteria*. Contrary to our initial expectations, we found that both  
396 protein families display limited taxonomic distribution. SbmA/BacA orthologs were identified only in  
397 the phylum *Pseudomonadales*, with ~89% of the identified BacA proteins being encoded by species of  
398 the classes *Alphaproteobacteria* and *Gammaproteobacteria*. A majority of the identified BclA proteins  
399 were also found in species of the phylum *Pseudomonadales* with a bias towards the *Betaproteobacteria*;  
400 however, BclA proteins were also common in the phylum *Cyanobacteriota*, the class *Negativicutes*  
401 (phylum *Bacillota*), and the order *Mycobacteriales* (phylum *Actinomycetota*). Interestingly, many of the  
402 clades enriched for species encoding SbmA/BacA or BclA orthologs also include many species known  
403 to interact with eukaryotic hosts in mutualistic or pathogenic interactions.

404 Forty-five of the 55 species (~82%) of the alphaproteobacterial order *Hyphomicrobiales* encode  
405 SbmA/BacA and/or BclA; this increases to 45 of 50 species (90%) when excluding the deep-branching

406 *Hyphomicrobiales* lineages. This order accounts for ~79% of the alphaproteobacterial species encoding  
407 SbmA/BacA and/or BclA orthologs. Many members of the order *Hyphomicrobiales* are notable for their  
408 ability to interact with eukaryotic hosts. All alpha-rhizobia belong to the order *Hyphomicrobiales*, which  
409 also encompasses several plant and mammalian pathogens like *Agrobacterium* and *Brucella*,  
410 respectively (39). Similarly, ~75% of the gammaproteobacterial BacA and BclA proteins are encoded  
411 by species in the orders *Enterobacteriales* and *Pasteurellales*, in which 34 of 47 (~72%; increasing to 81%  
412 when excluding a monophyletic group of five obligate endosymbionts) and 10 of 16 (~62.5%) species  
413 encode BacA/BclA, respectively. The order *Enterobacteriales* is well-known for including many plant  
414 (e.g., *Dickeya*, *Pantoea*) and animal/human (e.g., *Klebsiella*, *Yersinia*) pathogens (40). Likewise, the order  
415 *Pasteurellales* encompasses several animal/human pathogens (e.g., *Haemophilus*, *Pasteurella*) (41). In  
416 the class *Betaproteobacteria*, BclA and SbmA/BacA were significantly more common in the orders  
417 *Burkholderiales* and *Neisseriales* compared to the orders *Nitrosomonadales* and *Rhodocyclales*. The  
418 order *Burkholderiales* encompass all known beta-rhizobia as well as insect gut symbionts (e.g.,  
419 *Caballeronia*) and plant (e.g., *Ralstonia*) and animal/human (e.g., *Burkholderia*) pathogens (42, 43). The  
420 order *Neisseriales* encompasses many mammalian commensals but also some human pathogens (e.g.,  
421 *Neisseria*) (44).

422 The phylum *Cyanobacteria* is the largest clade of organisms encoding BclA proteins outside of  
423 the phylum *Pseudomonadales*. To our knowledge, cyanobacteria are not pathogenic. However, many  
424 can form beneficial associations with diverse hosts, such as the nitrogen-fixing symbiosis between  
425 *Nostoc* and plants (45), the mutualistic relationship with fungi (forming lichens), and with sponges (46).  
426 The order *Mycobacteriales* includes important human and plant pathogens (e.g., *Mycobacterium*,  
427 *Rhodococcoides*) (47), and opportunistic pathogens (e.g., *Mycolicibacterium*) (48). The class *Negativicutes*  
428 is poorly studied despite its peculiar nature, as these *Firmicutes* possess an outer membrane and a  
429 LPS (49). Nevertheless, this class is a common component of eukaryotic microbiomes and can cause  
430 human disease, including meningitis (50).

431 The observation that most taxonomic clades enriched for species encoding SbmA/BacA or BclA  
432 also contain many mutualistic and/or pathogenic organisms may suggest that eukaryotic host  
433 interaction is a driver of SbmA/BacA and BclA maintenance in these lineages. However, the data also  
434 suggests that these protein families may pre-date these species interactions. Assuming that  
435 SbmA/BacA was acquired by the common ancestor of the BacA-containing subclade of the order  
436 *Hyphomicrobiales*, the SbmA/BacA protein family potentially evolved in this lineage over 500 million  
437 years ago (51), which predates the evolution of legumes that are estimated to have evolved around

438 60 million years ago (52). Thus, SbmA/BacA could not have evolved in this lineage as a response to  
439 legume symbiosis. Accordingly, rhizobia that do not face NCR peptides in their legume host have a  
440 SbmA/BacA protein that can transport peptides (24) while not being able to complement the symbiotic  
441 defect of a *S. meliloti bacA* mutant (28). Rather, we hypothesize that SbmA/BacA originally evolved to  
442 fulfil another role and was subsequently co-opted to support legume symbiosis in rhizobia. Likewise,  
443 we hypothesize that BclA already existed in the *Bradyrhizobium* lineage prior to the evolution of  
444 legume symbiosis, and that this protein was independently co-opted for legume symbiosis in these  
445 organisms, mimicking the convergent evolution of NCR peptides in the IRLC and Dalbergioid legume  
446 families (12). Another role for SbmA/BacA proteins that may predate the evolution of symbiosis is inter  
447 and intraspecific competition, as highlighted by a study of phazolicin-producing rhizobia (53).  
448 Phazolicin is a narrow-spectrum antibiotic peptide that is produced by some rhizobial strains and that  
449 can kill other rhizobia after being imported by BacA and YejABEF transporters (53). More broadly, we  
450 hypothesize that BacA and BclA proteins did not evolve specifically for host interaction but rather were  
451 repeatedly co-opted to help bacteria survive exposure to host encoded antimicrobial peptides (e.g.,  
452 NCR peptides in rhizobium-legume symbioses, and mammalian immune antimicrobial peptides  
453 during infection).

454

#### 455 **Functional characterization of SbmA/BacA and BclA protein families**

456 The abilities of several newly identified BclA proteins to complement the phenotypes of a *S. meliloti*  
457  $\Delta bacA$  mutant were tested to validate that these proteins were correctly annotated. *S. meliloti bacA*  
458 null mutants display increased gentamicin resistance compared to the wild type (20). Eight of the nine  
459 synthesized genes at least partially complemented the gentamicin resistance phenotype of a *S. meliloti*  
460  $\Delta bacA$  mutant, suggesting these eight proteins were expressed and at least partially functional in *S.*  
461 *meliloti*. Interestingly, even the gene encoding an ExsE ortholog partially complemented the  
462 gentamicin resistance phenotype, indicating that gentamicin transport is not specific to SbmA/BacA  
463 and BclA proteins but is a general property of these and related protein families. Gentamicin  
464 sensitivity assays are commonly used to characterize the function of rhizobial *bacA* orthologs and  
465 rhizobial *bacA* mutant alleles generated through site-directed mutagenesis (26). Although these assays  
466 are useful to identify null phenotypes, our results show that they do not probe a function unique to  
467 SbmA/BacA or BclA proteins and thus have limited value as a proxy to peptide transport or host  
468 interaction assays.

469 In addition to showing increased resistance to gentamicin, *S. meliloti*  $\Delta bacA$  mutants show

470 increased sensitivity to NCR peptides (17, 26). As the antimicrobial activity of NCR peptides is a result  
471 of their interaction with the cell envelope, it is thought that SbmA/BacA and BclA proteins provide  
472 resistance to NCR peptides by moving the peptides away from the cell envelope and into the cell (17,  
473 28). SbmA/BacA and BclA proteins have also been shown to transport other antimicrobial peptides,  
474 including mammalian antimicrobial peptides such as Bac7 (19, 33, 36, 54). As expected, only the  
475 proteins annotated as BclA were capable of effectively complementing the sensitivity of *S. meliloti*  
476  $\Delta bacA$  and *S. meliloti*  $\Delta bacA \Omega yejA$  mutants to the NCR peptide NCR247. Of the six newly-identified BclA  
477 proteins that were tested, four repeatedly demonstrated good levels of complementation; these  
478 proteins were from *P. naphthalenivorans* (class *Betaproteobacteria*), *S. dextrinosolvens* (class  
479 *Gammaproteobacteria*), *S. elongatus* (phylum *Cyanobacteriota*), and *C. aponinum* (phylum  
480 *Cyanobacteriota*). The other two, from *M. anaerophila* (class *Negativicutes*) and *B. psittacipulmonis* (class  
481 *Betaproteobacteria*), showed weak and variable or little to no complementation, respectively. However,  
482 there are thousands of distinct NCR peptides encoded across the legume family (11), and thus the  
483 inability of a transporter to transport NCR247 does not mean that it is unable to transport other NCR  
484 peptides, or mammalian antimicrobial peptides. Indeed, *S. meliloti* *yjA* mutants show increased  
485 sensitivity to the peptide NCR280 but not NCR247 (30). Regardless, these results support that the  
486 ability to transport antimicrobial peptides, including NCR peptides, is a general property of bacterial  
487 SbmA-like proteins.

488

## 489 **Conclusions**

490 In summary, we identified 208 bacterial species encoding SbmA/BacA or BclA. These species were not  
491 equally distributed across the domain *Bacteria*; instead, SbmA/BacA proteins were found only in the  
492 phylum *Pseudomonadota*, while BclA proteins were primarily found within a subset of families across  
493 four phyla. Our analyses suggest that the SbmA/BacA and BclA protein families arose independently  
494 and that their functional similarity is a result of convergent evolution rather than shared ancestry. Our  
495 data also support the hypothesis that SbmA/BacA and BclA proteins have been repeatedly co-opted  
496 to facilitate both mutualistic and pathogenic associations with eukaryotic hosts by allowing bacteria  
497 to cope with host-encoded antimicrobial peptides.

498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529

## MATERIALS AND METHODS

### Bacterial strains and growth conditions

The bacterial strains used in this study are listed in **Table S2**. *E. coli* strains were cultured at 37 °C using Lysogeny Broth (LB; 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl). *S. meliloti* strains were grown at 28 °C using either LBmc (LB supplemented with 2.5 mM CaCl<sub>2</sub> and 2.5 mM MgSO<sub>4</sub>), YEB (0.5% beef extract, 0.1% yeast extract, 0.5% peptone, 0.5% sucrose, 0.04% MgSO<sub>4</sub> 7H<sub>2</sub>O, pH 7.5), or MM9 minimal medium (2% MOPS-KOH, 1.92% NH<sub>4</sub>Cl, 0.35% NaCl, 0.2% KH<sub>2</sub>PO<sub>4</sub>, 0.2% MgSO<sub>4</sub>, 0.05% CaCl<sub>2</sub>, 0.05% Biotin, 0.0004% CoCl<sub>2</sub>, 0.38% FeCl<sub>3</sub>, 1% Glucose, 1% Na<sub>2</sub>-succinate). Antibiotics were added as appropriate and included: ampicillin (Amp; 100 µg/mL), kanamycin (Km; 100 µg/mL), streptomycin (Sm; 200 or 500 µg/mL), spectinomycin (Sp; 50 µg/mL), and tetracycline (Tc; 5 µg/mL). Antibiotic concentrations were generally halved for liquid cultures.

### Cloning of *bacA*, *bclA*, and *exsE* homologs

Ten vectors encoding putative *bacA*, *bclA*, or *exsE* genes, codon optimized for *S. meliloti* 1021 and flanked by XbaI and BamHI recognition sites, were produced by Twist Biosciences (**Table S2, Dataset S1**). Each gene was PCR amplified from the plasmids using Q5 polymerase (New England Biolabs; NEB) with the primers 5'-GAAGTGCCATTCCGCCTGACC and 5'-CACTGAGCCTCCACCTAGCC. The resulting amplicons were individually digested with XbaI/BamHI and ligated into XbaI/BamHI-digested expression vector pRF771 (55). Plasmids were sequence verified via Illumina sequencing (151 bp paired-end reads) at SeqCenter (Pittsburg, PA, USA), after which reads were aligned to the expected template sequences using bowtie2 version 2.5.0 (56) and alignments visualized using the Integrative Genomics Viewer version 2.12.3 (57).

### Transfer of plasmids to *S. meliloti*

All plasmids of interest were transferred to a *S. meliloti*  $\Delta bacA$  mutant via triparental matings using the helper strains *E. coli* MT616 or *E. coli* HB101, as described previously (37, 58). Transconjugants were recovered through plating of mating spots on LBmc Sm<sup>200</sup> Tc or YEB Sm<sup>500</sup> Tc plates. Likewise, plasmids were transferred to a *S. meliloti*  $\Delta bacA$   $\Omega yejA$  double mutant via triparental mating as described previously (37), with transconjugants recovered on YEB Sm<sup>500</sup> Tc Km Sp plates. All transconjugants were streak purified three times prior to use.

530 **Gentamicin sensitivity assays**

531 Gentamicin sensitivity assays were performed largely as described previously (28). Briefly, overnight  
532 cultures of *S. meliloti*, grown in LBmc Sm<sup>100</sup> Tc, were washed and resuspended in LBmc to an optical  
533 density at 600 nm (OD<sub>600</sub>) of 1.0. Ten µL aliquots of the cell suspensions were added to triplicate wells  
534 of a 96-well plate and mixed with 190 µL of LBmc with or without 20 µg/mL of gentamicin (Gm). A Gm  
535 concentration of 20 µg/mL was chosen for the assays based on preliminary sensitivity assays (**Figure**  
536 **S4**). Plates were tape-closed to prevent evaporation and then incubated at 30°C with maximal shaking  
537 in a BioTek Synergy H1 plate reader for 24 hours. OD<sub>600</sub> measurements were collected every 15  
538 minutes using the Gen5 software (Agilent Technologies).

539

540 **NCR247 sensitivity assays**

541 NCR sensitivity assays were performed largely as described previously (30). Briefly, overnight cultures  
542 of *S. meliloti*, grown in MM9 minimal media, were washed and resuspended in MM9 to an OD<sub>600</sub> of  
543 1.0. Cell suspensions were then diluted to an OD<sub>600</sub> of 0.05 and 145 µL transferred to the wells of 96-  
544 well plates and mixed with 5 µL of an NCR247 solution to reach final concentrations of 50, 25, 12.5,  
545 6.25, 3.125, and 0 µg/mL of NCR247. Plates were incubated at 28°C with shaking (180 rpm) in a Tecan  
546 Spark plate reader for 72 hours, and OD<sub>600</sub> measurements were taken every 30 minutes and processed  
547 using the SparkControl software (Tecan).

548

549 **Plant symbiotic assays**

550 Seeds of *M. sativa* cv. Algonquin (alfalfa) and *M. officinalis* (yellow-blossom sweet clover) (Speare Seeds  
551 Limited; Harriston, Ontario, Canada) were surface-sterilized and germinated on water agar plates for  
552 two nights in the dark, as described previously (13). Leonard assemblies were prepared as described  
553 before (13), with a 1:1 (w/w) mixture of vermiculite and silica sand in the top pot, 250 mL Jensen's  
554 medium (59) in the bottom pot, and a cotton wick connecting the pots, and then autoclaved. Five  
555 seedlings were sown per pot, and assemblies were incubated for two nights. Assemblies were next  
556 inoculated in triplicate with 1 × 10<sup>8</sup> CFU of *S. meliloti* per assembly. Plants were grown in a Conviron  
557 growth chamber with an 18-hour photoperiod, 300 µmol/s of light, 21 °C daytime temperature, and  
558 17 °C nighttime temperature. After 30 days, plant shoots were collected and dried at 60 °C for six  
559 nights prior to weighing.

560

561



## 562 **Phylogenetic analysis of BacA and BclA proteins**

563 GenBank files corresponding to 3498 RefSeq bacterial genomes with 'complete' genome assemblies  
564 were downloaded from the National Center for Biotechnology Information (NCBI) Genome Database.  
565 A subset of the genomes was prepared by collecting genomes from one representative genome per  
566 genus, using the genome from the first species per genus when sorted alphabetically. The  
567 phylogenetic analyses were then repeated twice: once using all 3498 RefSeq bacterial genomes and  
568 once using the reduced set of 1255 genomes (**Dataset S2**). As the results were similar, we only present  
569 results generated using the reduced dataset.

570 BacA, BclA, and related proteins were extracted from the bacterial proteomes using a modified  
571 version of an existing in-house pipeline (60). The seed alignment of the SbmA/BacA-like family,  
572 consisting of eight sequences, was downloaded from PFAM (PF05992), and a hidden Markov model  
573 (HMM) built using the hmmbuild function of HMMER version 3.3 (61). Separately a HMM database was  
574 built by combining (i) the complete PFAM-A version 31.0 HMM database, (ii) the complete TIGERFAM  
575 version 15.0 HMM database, (iii) HMMs built from the seed alignments of PRK11098 (105 sequences  
576 in the seed alignment) and COG1133 (nine sequences in the seed alignment) downloaded from NCBI's  
577 Conserved Domain Database, and (iv) HMMs built for each of the BacA (15 sequences), BclA (5  
578 sequences), *Mycobacterium* BacA (10 sequences), ExsE (6 sequences), and *Bradyrhizobium* homologous  
579 clade (7 sequences) proteins used in the phylogenetic analysis of Guefrachi et al. 2015 (36). Next, the  
580 hmmsearch function of HMMER was used to search all bacterial proteomes using the PF05992  
581 (SbmA/BacA-like family) HMM. All hmmsearch hits were then scanned against the full HMM database  
582 using the hmmscan function of HMMER. Each protein was annotated according to the top-scoring  
583 HMM from this search.

584 Proteins annotated as BacA, BclA, *Mycobacterium* BacA, ExsE, or *Bradyrhizobium* homologous  
585 clade were extracted and aligned using Clustal Omega version 1.2.4 (62), hmalign from HMMER (61),  
586 and MAFFT version 7.453 (63), and alignment quality assessed with T-COFFEE version 13.45 (64). Poor  
587 quality regions of the best scoring alignment (Clustal Omega) were removed using trimAl version 1.4  
588 with the automated1 option (65), and then used as input for maximum likelihood phylogeny inference  
589 using IQ-TREE2 version 2.2.0 (66) with the LG+F+I+R9 model. The LG+F+I+R9 model was used as it  
590 was identified as the best-scoring model by the IQ-TREE2 implementation of ModelFinder (67) based  
591 on Bayesian information criterion (BIC), with model search limited to the LG, WAG, JTT, Q.pfam,  
592 JTTDCMut, DCMut, VT, PMB, BLOSUM62, and Dayhoff models. Branch supports were assessed in IQ-  
593 TREE using Shimodaira-Hasegawa-like approximate likelihood ratio test (SH-aLRT) (68) and an ultrafast



594 bootstrap analysis, with both metrics calculated from 1000 replicates. All phylogenies created in this  
595 study were visualized with the iTOL web server (69).

596

### 597 **Sequence similarity network analysis**

598 A sequence similarity network (SSN) was constructed for the 366 proteins identified using the HMM  
599 approach described above. The SSN was constructed using the online Enzyme Function Initiative's  
600 Enzyme Similarity Tool (EFI-EST; <https://efi.igb.illinois.edu/efi-est/>) (70, 71) with an alignment score  
601 threshold of 115, corresponding to an approximate sequence ID  $\geq$  35%. The resulting network was  
602 visualized using Cytoscape version 3.10.1 (72).

603

### 604 **Multilocus sequence analysis**

605 A bacterial species phylogeny was produced for the 1,253 representative bacterial species using an  
606 adaptation of an existing in-house pipeline (60); two of the 1,255 downloaded genomes were excluded  
607 as they encoded none of the marker genes. First, orthologs of 31 highly-conserved, single-copy  
608 proteins (DnaG, Frr, InfC, NusA, Pkg, PyrG, RplA, RplB, RplC, RplD, RplE, RplF, RplK, RplL, RplM, RplN,  
609 RplP, RplS, RplT, RpmA, RpoB, RpsB, RpsC, RpsE, RpsI, RpsJ, RpsK, RpsM, RpsS, SmpB, Tsf) were  
610 identified in the 1,253 proteomes using the AMPHORA2 pipeline (73). Each group of orthologs was  
611 individually aligned using MAFFT (63) and trimmed using trimAl with the option (65). The protein  
612 alignments were then concatenated and used as input for ModelFinder as implemented in IQ-TREE2,  
613 and the best scoring model was identified based on BIC. IQ-TREE2 was then used to infer a maximum  
614 likelihood phylogeny from the concatenated alignment using the LG+I+I+R10 model. Branch supports  
615 were assessed in IQ-TREE using the Shimodaira-Hasegawa-like approximate likelihood ratio test (SH-  
616 aLRT) [22] and ultrafast jackknife analysis with a subsampling proportion of 40%, with both metrics  
617 calculated from 1000 replicates.

618

### 619 **Data availability**

620 All genome sequences used in this work were previously published, and the assembly accessions are  
621 provided in **Dataset S2**. Likewise, all protein sequences included in Figure 1 are provided in **Dataset**  
622 **S1**. Newick formatted phylogenies used to create Figures 1 and 4 are available through GitHub  
623 ([https://github.com/amira-boukh/Sbma\\_BacA\\_phylogenetic\\_distribution](https://github.com/amira-boukh/Sbma_BacA_phylogenetic_distribution)). All code to repeat the  
624 analyses in this study is also available through GitHub ([https://github.com/amira-](https://github.com/amira-boukh/Sbma_BacA_phylogenetic_distribution)  
625 [boukh/Sbma\\_BacA\\_phylogenetic\\_distribution](https://github.com/amira-boukh/Sbma_BacA_phylogenetic_distribution)).

626

## ACKNOWLEDGEMENTS

627 This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC)  
628 Discovery Grants to G.C.D. and G.W.H.. N.T.S. was supported, in part, by an R. S. McLaughlin Fellowship  
629 from Queen's University and a Wicked Ideas grant from Queen's University to G.W.H and G.C.D.. A.B.  
630 benefited from a Ph.D. contract in the frame of the CNRS 80|PRIME – 2021 program and was partially  
631 supported by a Mitacs Globalink Research Award. B.A. benefited from a French State grant (Saclay  
632 Plant Sciences, reference n° ANR-17-EUR-0007, EUR SPS-GSR) under a France 2030 program (reference  
633 n° ANR-11-IDEX-0003).

634

635

## REFERENCES

- 636 1. Poole P, Ramachandran V, Terpolilli J. 2018. Rhizobia: From saprophytes to endosymbionts.  
637 Nature Reviews Microbiology 16:291–303.
- 638 2. Menegat S, Ledo A, Tirado R. 2022. Greenhouse gas emissions from global production and use  
639 of nitrogen synthetic fertilisers in agriculture. Sci Rep 12:14490.
- 640 3. Peoples MB, Brockwell J, Herridge DF, Rochester IJ, Alves BJR, Urquiaga S, Boddey RM, Dakora  
641 FD, Bhattarai S, Maskey SL, Sampet C, Rerkasem B, Khan DF, Hauggaard-Nielsen H, Jensen ES.  
642 2009. The contributions of nitrogen-fixing crop legumes to the productivity of agricultural  
643 systems. Symbiosis 48:1–17.
- 644 4. Alunni B, Gourion B. 2016. Terminal bacteroid differentiation in the legume–rhizobium  
645 symbiosis: nodule-specific cysteine-rich peptides and beyond. New Phytologist 211:411–417.
- 646 5. Haag AF, Mergaert P. 2020. Terminal bacteroid differentiation in the *Medicago–Rhizobium*  
647 interaction – a tug of war between plant and bacteria, p. 600–616. *In* The Model Legume *Medicago*  
648 *truncatula*. John Wiley & Sons, Ltd.
- 649 6. Mergaert P, Uchiumi T, Alunni B, Evanno G, Cheron A, Catrice O, Mausset A-E, Barloy-Hubler F,  
650 Galibert F, Kondorosi A, Kondorosi E. 2006. Eukaryotic control on bacterial cell cycle and  
651 differentiation in the Rhizobium–legume symbiosis. Proceedings of the National Academy of  
652 Sciences 103:5230–5235.
- 653 7. Roux B, Rodde N, Jardinaud M-F, Timmers T, Sauviac L, Cottret L, Carrère S, Sallet E, Courcelle E,  
654 Moreau S, Debelle F, Capela D, de Carvalho-Niebel F, Gouzy J, Bruand C, Gamas P. 2014. An  
655 integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-  
656 capture microdissection coupled to RNA sequencing. Plant J 77:817–837.
- 657 8. diCenzo GC, Tesi M, Pfau T, Mengoni A, Fondi M. 2020. Genome-scale metabolic reconstruction

- 658 of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. 1. Nat Commun  
659 11:2574.
- 660 9. Van De Velde W, Zehirov G, Szatmari A, Debreczeny M, Ishihara H, Kevei Z, Farkas A, Mikulass K,  
661 Nagy A, Tiricz H, Satiat-Jeunemaître B, Alunni B, Bourge M, Kucho KI, Abe M, Kereszt A, Maroti G,  
662 Uchiumi T, Kondorosi E, Mergaert P. 2010. Plant peptides govern terminal differentiation of  
663 bacteria in symbiosis. Science 327:1122–1126.
- 664 10. Kereszt A, Mergaert P, Montiel J, Endre G, Kondorosi É. 2018. Impact of Plant Peptides on  
665 Symbiotic Nodule Development and Functioning. Frontiers in Plant Science 9:1026.
- 666 11. Montiel J, Downie JA, Farkas A, Bihari P, Herczeg R, Bálint B, Mergaert P, Kereszt A, Kondorosi É.  
667 2017. Morphotype of bacteroids in different legumes correlates with the number and type of  
668 symbiotic NCR peptides. Proceedings of the National Academy of Sciences 114:5041–5046.
- 669 12. Czernic P, Gully D, Cartieaux F, Moulin L, Guefrachi I, Patrel D, Pierre O, Fardoux J, Chaintreuil C,  
670 Nguyen P, Gressent F, Da Silva C, Poulain J, Wincker P, Rofidal V, Hem S, Barrière Q, Arrighi J-F,  
671 Mergaert P, Giraud E. 2015. Convergent Evolution of Endosymbiont Differentiation in Dalbergioid  
672 and Inverted Repeat-Lacking Clade Legumes Mediated by Nodule-Specific Cysteine-Rich  
673 Peptides. Plant Physiol 169:1254–1265.
- 674 13. Huang R, Snedden WA, diCenzo GC. 2022. Reference nodule transcriptomes for *Melilotus*  
675 *officinalis* and *Medicago sativa* cv. Algonquin. Plant Direct 6:e408.
- 676 14. Szerencsés B, Gácsér A, Endre G, Domonkos I, Tiricz H, Vágvölgyi C, Szolomajer J, Howan DHO,  
677 Tóth GK, Pfeiffer I, Kondorosi É. 2021. Symbiotic NCR Peptide Fragments Affect the Viability,  
678 Morphology and Biofilm Formation of *Candida* Species. Int J Mol Sci 22:3666.
- 679 15. Tiricz H, Szücs A, Farkas A, Pap B, Lima RM, Maróti G, Kondorosi É, Kereszt A. 2013. Antimicrobial  
680 nodule-specific cysteine-rich peptides induce membrane depolarization-associated changes in  
681 the transcriptome of *Sinorhizobium meliloti*. Applied and Environmental Microbiology 79:6737–  
682 6746.
- 683 16. Penterman J, Abo RP, De Nisco NJ, Arnold MFF, Longhi R, Zanda M, Walker GC. 2014. Host plant  
684 peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell cycle during  
685 symbiosis. Proceedings of the National Academy of Sciences 111:3561–3566.
- 686 17. Haag AF, Baloban M, Sani M, Kerscher B, Pierre O, Farkas A, Longhi R, Boncompagni E, Hérouart  
687 D, Dall'Angelo S, Kondorosi E, Zanda M, Mergaert P, Ferguson GP. 2011. Protection of  
688 *Sinorhizobium* against host cysteine-rich antimicrobial peptides is critical for symbiosis. PLoS  
689 Biology 9:e1001169.

- 690 18. Ghilarov D, Inaba-Inoue S, Stepien P, Qu F, Michalczyk E, Pakosz Z, Nomura N, Ogasawara S,  
691 Walker GC, Rebuffat S, Iwata S, Heddle JG, Beis K. 2021. Molecular mechanism of SbmA, a  
692 promiscuous transporter exploited by antimicrobial peptides. *Science Advances* 7:eabj5363.
- 693 19. Runti G, Lopez Ruiz M del C, Stoilova T, Hussain R, Jennions M, Choudhury HG, Benincasa M,  
694 Gennaro R, Beis K, Scocchi M. 2013. Functional characterization of SbmA, a bacterial inner  
695 membrane transporter required for importing the antimicrobial peptide Bac7. *J Bacteriol*  
696 195:5343–5351.
- 697 20. Ichige A, Walker GC. 1997. Genetic analysis of the *Rhizobium meliloti* *bacA* gene: Functional  
698 interchangeability with the *Escherichia coli* *sbmA* gene and phenotypes of mutants. *J Bacteriol*  
699 179:209-216.
- 700 21. Karunakaran R, Haag AF, East AK, Ramachandran VK, Prell J, James EK, Scocchi M, Ferguson GP,  
701 Poole PS. 2010. BacA Is Essential for Bacteroid Development in Nodules of Galeoid, but not  
702 Phaseoloid, Legumes. *J Bacteriol* 192:2920–2928.
- 703 22. Tan X-J, Cheng Y, Li Y-X, Li Y-G, Zhou J-C. 2009. BacA is indispensable for successful  
704 *Mesorhizobium-Astragalus* symbiosis. *Appl Microbiol Biotechnol* 84:519–526.
- 705 23. Maruya J, Saeki K. 2010. The *bacA* Gene Homolog, *mlr7400*, in *Mesorhizobium loti* MAFF303099 is  
706 dispensable for symbiosis with *Lotus japonicus* but partially capable of supporting the symbiotic  
707 function of *bacA* in *Sinorhizobium meliloti*. *Plant and Cell Physiology* 51:1443–1452.
- 708 24. Ardisson S, Kobayashi H, Kambara K, Rummel C, Noel KD, Walker GC, Broughton WJ, Deakin WJ.  
709 2011. Role of BacA in lipopolysaccharide synthesis, peptide transport, and nodulation by  
710 *Rhizobium* sp. strain NGR234. *J Bacteriol* 193:2218–2228.
- 711 25. Ferguson GP, Datta A, Carlson RW, Walker GC. 2005. Importance of unusually modified lipid A in  
712 *Sinorhizobium* stress resistance and legume symbiosis. *Mol Microbiol* 56:68–80.
- 713 26. LeVier K, Walker GC. 2001. Genetic analysis of the *Sinorhizobium meliloti* BacA protein: Differential  
714 effects of mutations on phenotypes. *J Bacteriol* 183:6444–6453.
- 715 27. Glazebrook J, Ichige A, Walker GC. 1993. A *Rhizobium meliloti* homolog of the *Escherichia coli*  
716 peptide-antibiotic transport protein SbmA is essential for bacteroid development. *Genes and*  
717 *Development* 7:1485–1497.
- 718 28. DiCenzo GC, Zamani M, Ludwig HN, Finan TM. 2017. Heterologous complementation reveals a  
719 specialized activity for BacA in the *Medicago-Sinorhizobium meliloti* symbiosis. *Molecular Plant-*  
720 *Microbe Interactions* 30:312–324.
- 721 29. Farkas A, Maróti G, Dürgo H, Györgypál Z, Lima RM, Medzihradszky KF, Kereszt A, Mergaert P,

- 722 Kondorosi É. 2014. *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid  
723 differentiation through multiple mechanisms. Proceedings of the National Academy of Sciences.  
724 111:5183–5188.
- 725 30. Nicoud Q, Barrière Q, Busset N, Dendene S, Travin D, Bourge M, Le Bars R, Boulogne C, Lacroël  
726 M, Jenei S, Kereszt A, Kondorosi E, Biondi EG, Timchenko T, Alunni B, Mergaert P. 2021.  
727 *Sinorhizobium meliloti* Functions Required for Resistance to Antimicrobial NCR Peptides and  
728 Bacteroid Differentiation. mBio 12:10.1128/mbio.00895-21.
- 729 31. Laviña M, Pugsley AP, Moreno F. 1986. Identification, mapping, cloning and characterization of a  
730 gene (*sbmA*) required for microcin B17 action on *Escherichia coli* K12. J Gen Microbiol 132:1685–  
731 1693.
- 732 32. LeVier K, Phillips RW, Grippe VK, Roop RM, Walker GC. 2000. Similar requirements of a plant  
733 symbiont and a mammalian pathogen for prolonged intracellular survival. Science 287:2492–  
734 2493.
- 735 33. Domenech P, Kobayashi H, Levier K, Walker GC, Iii CEB. 2009. BacA, an ABC Transporter Involved  
736 in Maintenance of Chronic Murine Infections with *Mycobacterium tuberculosis*. J Bacteriol  
737 191:477–485.
- 738 34. Li G, LTURNUS C, Ewers C, Wieler LH. 2005. Identification of Genes Required for Avian *Escherichia*  
739 *coli* Septicemia by Signature-Tagged Mutagenesis. Infection and Immunity 73:2818.
- 740 35. Wehmeier S, Arnold MFF, Marlow VL, Aouida M, Myka KK, Fletcher V, Benincasa M, Scocchi M,  
741 Ramotar D, Ferguson GP. 2010. Internalization of a thiazole-modified peptide in *Sinorhizobium*  
742 *meliloti* occurs by BacA-dependent and -independent mechanisms. Microbiology 156:2702–2713.
- 743 36. Guefrachi I, Pierre O, Timchenko T, Alunni B, Barrière Q, Czernic P, Villaécija-Aguilar JA, Verly C,  
744 Bourge M, Fardoux J, Mars M, Kondorosi E, Giraud E, Mergaert P. 2015. *Bradyrhizobium* BclA Is a  
745 Peptide Transporter Required for Bacterial Differentiation in Symbiosis with *Aeschynomene*  
746 Legumes. Molecular Plant-Microbe Interactions 28:1155–1166.
- 747 37. Barrière Q, Guefrachi I, Gully D, Lamouche F, Pierre O, Fardoux J, Chaintreuil C, Alunni B,  
748 Timchenko T, Giraud E, Mergaert P. 2017. Integrated roles of BclA and DD-carboxypeptidase 1 in  
749 *Bradyrhizobium* differentiation within NCR-producing and NCR-lacking root nodules. Sci Rep  
750 7:9063.
- 751 38. Arnold MFF, Haag AF, Capewell S, Boshoff HI, James EK, Donald RM, Mair I, Mitchell AM, Kerscher  
752 B, Mitchell TJ, Mergaert P, Barry IE, Scocchi M, Zanda M, Campopiano DJ, Ferguson GP. 2013.  
753 Partial complementation of *Sinorhizobium meliloti* *baca* mutant phenotypes by the

- 754 *Mycobacterium tuberculosis* BacA protein. J Bacteriol 195:389-398.
- 755 39. diCenzo GC, Yang Y, Young JPW, Kuzmanović N. 2023. Refining the taxonomy of the order  
756 *Hyphomicrobiales* (*Rhizobiales*) based on whole genome comparisons of over 130 genus type  
757 strains. bioRxiv <https://doi.org/10.1101/2023.11.15.567303>.
- 758 40. de la Maza LM, Pezzlo MT, Bittencourt CE, Peterson EM. 2020. Introduction to *Enterobacterales*,  
759 p. 91–102. In Color Atlas of Medical Bacteriology. John Wiley & Sons, Ltd.
- 760 41. Garrity GM, Bell JA, Lilburn T. 2005. *Pasteurellales* ord. nov., p. 850–912. In Brenner, DJ, Krieg, NR,  
761 Staley, JT, Garrity, GM, Boone, DR, De Vos, P, Goodfellow, M, Rainey, FA, Schleifer, K-H (eds.),  
762 Bergey's Manual® of Systematic Bacteriology: Volume Two The Proteobacteria Part B The  
763 Gammaproteobacteria. Springer US, Boston, MA.
- 764 42. Voronina OL, Kunda MS, Ryzhova NN, Aksenova EI, Semenov AN, Lasareva AV, Amelina EL,  
765 Chuchalin AG, Lunin VG, Gintsburg AL. 2015. The Variability of the Order Burkholderiales  
766 Representatives in the Healthcare Units. Biomed Res Int 2015:680210.
- 767 43. Dobritsa AP, Samadpour M. 2016. Transfer of eleven species of the genus Burkholderia to the  
768 genus Paraburkholderia and proposal of *Caballeronia* gen. nov. to accommodate twelve species  
769 of the genera Burkholderia and Paraburkholderia. International Journal of Systematic and  
770 Evolutionary Microbiology 66:2836–2846.
- 771 44. Chen S, Rudra B, Gupta RS. 2021. Phylogenomics and molecular signatures support division of  
772 the order *Neisseriales* into emended families *Neisseriaceae* and *Chromobacteriaceae* and three  
773 new families *Aquaspirillaceae* fam. nov., *Chitinibacteraceae* fam. nov., and *Leeiaceae* fam. nov.  
774 Systematic and Applied Microbiology 44:126251.
- 775 45. Bergman B, Johansson C, Söderbäck E. 1992. The *Nostoc-Gunnera* symbiosis. New Phytol  
776 122:379–400.
- 777 46. Mutalipassi M, Riccio G, Mazzella V, Galasso C, Somma E, Chiarore A, de Pascale D, Zupo V. 2021.  
778 Symbioses of Cyanobacteria in Marine Environments: Ecological Insights and Biotechnological  
779 Perspectives. Mar Drugs 19:227.
- 780 47. Val-Calvo J, Vázquez-Boland JA. 2023. *Mycobacteriales* taxonomy using network analysis-aided,  
781 context-uniform phylogenomic approach for non-subjective genus demarcation. mBio  
782 14:e02207-23.
- 783 48. Morgado S, Ramos N de V, Pereira BB do N, Freitas F, Fonseca ÉL da, Vicente AC. 2022. Multidrug-  
784 resistant *Mycolicibacterium fortuitum* infection in a companion cat (*Felis silvestris catus*) in Brazil.  
785 Access Microbiology 4:000317.



- 786 49. Antunes LC, Poppleton D, Klingl A, Criscuolo A, Dupuy B, Brochier-Armanet C, Beloin C, Gribaldo  
787 S. 2016. Phylogenomic analysis supports the ancestral presence of LPS-outer membranes in the  
788 Firmicutes. *eLife* 5:e14589.
- 789 50. Brown RF. 2016. Investigating the Evolutionary Origin and Cell Biology of Negativicutes.  
790 University of Warwick, Warwick.
- 791 51. Rahimlou S, Bahram M, Tedersoo L. 2021. Phylogenomics reveals the evolution of root  
792 nodulating alpha- and beta-Proteobacteria (rhizobia). *Microbiological Research* 250:126788.
- 793 52. Lavin M, Herendeen PS, Wojciechowski MF. 2005. Evolutionary Rates Analysis of *Leguminosae*  
794 Implicates a Rapid Diversification of Lineages during the Tertiary. *Systematic Biology* 54:575–594.
- 795 53. Travin DY, Jouan R, Vigouroux A, Inaba-Inoue S, Lachat J, Haq F, Timchenko T, Sutormin D, Dubiley  
796 S, Beis K, Moréra S, Severinov K, Mergaert P. 2023. Dual-Uptake Mode of the Antibiotic Phazolicin  
797 Prevents Resistance Acquisition by Gram-Negative Bacteria. *mBio* 14:e00217-23.
- 798 54. Marlow VL, Haag AF, Kobayashi H, Fletcher V, Scocchi M, Walker GC, Ferguson GP. 2009. Essential  
799 role for the BacA protein in the uptake of a truncated eukaryotic peptide in *Sinorhizobium meliloti*.  
800 *J Bacteriol* 191:1519–1527.
- 801 55. Wells DH, Long SR. 2002. The *Sinorhizobium meliloti* stringent response affects multiple aspects  
802 of symbiosis. *Molecular microbiology* 43:1115–1127.
- 803 56. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*  
804 2012 9:4 9:357–359.
- 805 57. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011.  
806 Integrative genomics viewer. *Nature Biotechnology* 2011 29:1 29:24–26.
- 807 58. Finan TM, Kunkel B, De Vos GF, Signer ER. 1986. Second symbiotic megaplasmid in *Rhizobium*  
808 *meliloti* carrying exopolysaccharide and thiamine synthesis genes. *J Bacteriol* 167:66.
- 809 59. Jensen HL, Jensen HL. 1942. Nitrogen fixation in leguminous plants. I. General characters of root-  
810 nodule bacteria isolated from species of *Medicago* and *Trifolium* in Australia. *Proceedings of the*  
811 *Linnean Society of New South Wales* 67:98–108.
- 812 60. diCenzo GC, Mengoni A, Perrin E. 2019. Chromids Aid Genome Expansion and Functional  
813 Diversification in the Family *Burkholderiaceae*. *Molecular Biology and Evolution* 36:562-574.
- 814 61. Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic and iterative HMM  
815 search procedure. *BMC Bioinformatics* 11:431.
- 816 62. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding  
817 J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple



- 818 sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
- 819 63. Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:  
820 Improvements in Performance and Usability. *Molecular Biology and Evolution* 30:772–780.
- 821 64. Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate  
822 multiple sequence alignment. *J Mol Biol* 302:205–217.
- 823 65. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment  
824 trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
- 825 66. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R.  
826 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic  
827 Era. *Molecular Biology and Evolution* 37:1530–1534.
- 828 67. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. 2017. ModelFinder: fast  
829 model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589.
- 830 68. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and  
831 methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML  
832 3.0. *Syst Biol* 59:307–321.
- 833 69. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree  
834 display and annotation. *Nucleic Acids Research* 49:W293–W296.
- 835 70. Oberg N, Zallot R, Gerlt JA. 2023. EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI)  
836 Web Resource for Genomic Enzymology Tools. *Journal of Molecular Biology* 435:168018.
- 837 71. Zallot R, Oberg N, Gerlt JA. 2019. The EFI Web Resource for Genomic Enzymology Tools:  
838 Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and  
839 Metabolic Pathways. *Biochemistry* 58:4169–4182.
- 840 72. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T.  
841 2003. Cytoscape: a software environment for integrated models of biomolecular interaction  
842 networks. *Genome Res* 13:2498–2504.
- 843 73. Wu M, Scott AJ. 2012. Phylogenomic analysis of bacterial and archaeal sequences with  
844 AMPHORA2. *Bioinformatics* 28:1033–1034.