Evolutionarily Driven Domain Swap Alters Sigma Factor Dependence in Bacterial Signaling System

Megan E. Garber[1,2], Vered Frank[2], Alexey E. Kazakov[3], Hanqiao Zhang[2,4], Matthew R. Incha[2,5], Lara Rajeev[2], Jay D. Keasling[2,4,6], Aindrila Mukhopadhyay*[1,2,3]

Author Affiliations:

1. University of California, Berkeley Department of Comparative Biochemistry
2. Lawrence Berkeley National Laboratory, Biological Systems and Engineering Division
3. Lawrence Berkeley National Laboratory, Environmental Genomics and Systems Biology Division
4. University of California, Berkeley Department of Bioengineering
5. University of California, Berkeley Department of Plant and Microbial Biology
6. University of California, Berkeley Department of Chemical and Biomolecular Engineering

* Correspondence: amukhopadhyay@lbl.gov

Abstract

Functional diversity in bacteria is introduced by lineage-specific expansion or horizontal gene transfer (HGT). Using modular bacterial signaling systems as a template, we experimentally validate domain swapping of modular proteins as an extension of the HGT model. We take a computational approach to explore the domain architecture of two-component systems (TCS) in select pseudomonads. We find a transcriptional effector domain swap that reconstructed a duplicated sigma54-dependent TCS to a sigma70-dependent TCS. Through functional genomics approaches, we determine that the implicated TCSs are involved in consumption of short-chain carboxylic acids. We verify the relationship between the domain-swapped TCSs utilizing a

mutational screen, in which we switch the specificity of the sigma70-dependent TCS output to the sigma54-dependent TCS input, and vice versa. Our findings suggest that this domain swap was maintained throughout α-, β-, γ- proteobacteria, thus domain swapping has potential to lead to fitness advantages and neofunctionalization.

Keywords

Domain swap; Two-component system; Response regulator; Bacterial evolution; Coevolution; Specificity; Insulation; Protein-protein interaction; Transcriptional regulation; Sigma factors; Sigma70; Sigma54; Protein-DNA interaction; DAP-seq; Functional genomics

Introduction

Bacteria use two-component systems (TCS) to sense and respond to signals in their environments. In a canonical TCS, a histidine kinase (HK) is auto-phosphorylated by ATP upon stimulation [1]. The phosphorylated HK can then engage in a phosphotransfer event to its cognate response regulator (RR) [2]. Although sequences encoding TCSs can be highly redundant in a single bacterial genome [3–7], biochemical and genetic evidence suggests that interactions between HKs and RRs are conserved between cognate pairs. Coevolutionary events of duplication and divergence have led to the apparent orthogonality of cognate pairs of HKs and RRs [8–16].

HKs and RRs are both modular proteins consisting of constant interacting domains, such as histidine-phosphotransfer (HPt), specifically HisKA, and catalytic ATPase (CA) domains in the HK and receiver (REC) domain in the RR (Figure 1a,b). Importantly, TCSs with HisKA domains are usually highly insulated [16], while HKs from other families, for instance with HisKA_2 or HWE domains, are otherwise promiscuous [17,18]. HKs and RRs are often found accessorized with a variety of N-terminal input or C-terminal effector domains [3,19–23]. The input domains in HKs typically encode for sensing functionality, whereas the appended domains in RRs confer output function. In the

46 context of the coevolution of TCS, the domains tethered to a pair of HKs and RRs are typically
47 carried through to descendants during the processes of lineage-specific evolution. However,
48 hypotheses of domain swapping in bacteria have emerged as a potential mechanism for the
49 diversification of TCSs [13,14,24,25].

50 In a domain swap the core parts of the TCS (HPt, CA, and REC domains) remain intact,
51 while the variable input or output domain is replaced by a different domain [24,25]. Domain swaps are
52 hypothesized to occur via horizontal gene transfer (HGT) and homologous recombination [25]. A
53 domain swap in a TCS can change the input or sensing parts of a HK, diversifying the molecular
54 inputs for the cascade [21,23], or it can alter the output domain of a RR, modifying the function of the
55 entire signalling cascade [19].

56 The output parts of RRs, the effector domains, dictate the final effect of a signaling
57 cascade. RRs can harbor output domains that lead to chemotactic responses [26,27], regulation of
58 small molecules such as cyclic nucleotides [28], or transcriptional regulation [4,19]. Transcriptional
59 effector domains (TEDs) interact primarily with DNA in a sequence-specific manner to regulate
60 the transcription of a gene. TEDs are diverse in structure, protein-DNA interaction, protein-protein
61 interaction, and function, and are therefore binned into separate sub-classes or families [19]. In this
62 study we predominantly observe TEDs associated with RRs that confer activation of genes with
63 sigma54 factors (AAA^+ tethered to HTH_8, NtrC family), and regulation of genes with sigma70
64 factors (GerE, NarL family or Trans_reg_C, OmpR family) (Figure 1c).

65 We take a computational approach to explore the domain architecture of RRs in a select
66 set of pseudomonads, with the goal of identifying evolutionarily driven domain swaps. All
67 computational predictions are then substantiated by original experimental results. Our findings
68 lead us to observe a domain swapping event that altered the sigma factor dependence of a cluster
69 of carboxylic acid responsive sigma54-dependent, NtrC-like RRs native to proteobacteria.

Results

Phylogenetic analysis of TCS domains from Pseudomonas RRs reveals TED swap

We identified all RRs in five representative pseudomonads: *P. aeruginosa* PAO1, *P. stutzeri* RCH2, *P. putida* KT2240, *P. fluorescens* FW300-N2E2, and *P. fluorescens* FW300-N2C3. We subdivided these RRs by their output domains, taking into account only RRs with TEDs (Supplementary Table 1). The phylogenetic tree of the REC domains importantly agrees with a previously published tree of *P. aeruginosa* PAO1 REC domains [29]. Our results demonstrate that the REC domains of RRs do not clade by species, but instead clade by TEDs (Figure 2, left panel). This observation is consistent with the current understanding how highly redundant proteins evolve in bacteria [16,30]. A domain swap can be hypothesized when RRs of one family cluster with RRs of another. Such an event can be observed within the AAA+-HTH_8/NtrC-like cluster (red and green) of the REC phylogenetic tree, where REC domains with an alternative TED, GerE/NarL-like (blue) are present (Figure 2, left panel).

To test a coevolutionary hypothesis between HKs and RRs, we applied the same phylogenetic strategy to sequences of HPt-CA (HisKA-HATPase_C) domains of all of the HKs in the same *Pseudomonas* strains. We observed that the leaves of the HPt-CA tree predominantly cluster by the domain architecture of their cognate RRs (Figure 2, right panel). We noted that the cognate HKs for the major clade of NarL-like RRs are hybrid HKs (grey bars), whereas the cognate HKs of the NarL-like RRs within the swapped cluster are more similar in domain architecture to the Ntr-like HKs. These results suggest that our initial observation and hypothesis of a functional domain swap was probable.

We next wanted to determine the evolutionary distance of the domain swap, asking whether it was present in pseudomonads only, or all proteobacteria. Using representatives from α-, β-, γ-, and δ- proteobacteria, selected for their high proportions of NtrC-like and NarL-like RRs, we were able to recapitulate the results of the *Pseudomonas* REC domain-based phylogenetic

95 tree. The proteobacteria REC phylogenetic tree reveals that the swap is found in the select strains
96 of α-, β-, and γ- proteobacteria, but not observed in the selected δ-proteobacteria (Supplementary
97 Figure 1). This discrepancy could be the result of gene loss in the selected representatives, or it
98 could indicate that the swapped clade never existed in δ-proteobacteria. Notably, in the new view
99 of the tree of life [31], it is hypothesized that δ-proteobacteria are more distantly related to the rest
100 of the proteobacterial clade than previously thought. Together these results lead to the hypothesis
101 that a domain swap likely occurred at the cuspis of the proteobacteria (α-, β-, γ-) clade.
102

103 *Determining the Functional Similarity of the RRs in the Swapped Cluster*

104 While the phylogenetic analysis can be used to hypothesize a proteobacteria-specific
105 domain swap, it does not inform on the functionality of the representatives. Closer observation of
106 the swapped cluster reveals that there are four subclusters, three of which harbor AAA+ domains
107 (NtrC-like/sigma54-dependent) and one of which harbors a GerE domain (NarL-like/sigma70-
108 dependent) (Figure 2 (close-up panel)). We hypothesized that if the domain swap is real, then the
109 signal cascades of the representatives within the clade should have similar functional inputs or
110 outputs. We focused our efforts on the functions of the TCSs in *P. putida* KT2440, because it is a
111 well studied model strain with a readily available genetic toolkit to enable follow-up and validation
112 studies.

113 Because the domain swap occurs within the NtrC-like cluster, we were able to hypothesize
114 that the signaling cascades were likely involved in nutrient uptake of carbon or nitrogen sources
115 [32–35]. We also used the annotations of the representatives within the swapped cluster that had been
116 previously identified as amino acid or di-carboxylic acid sensors [33,36–38]. However the functions of
117 the NarL-like representatives were unknown. We therefore applied high-throughput functional
118 genomics to explore the input signals and output genes of all of the relevant TCSs represented
119 within the cluster.

120 We applied an automated pipeline of DNA-affinity purification tethered to next generation
121 sequencing (DAP-seq), commonly used to biochemically determine genomic sites of protein-DNA
122 interaction for transcription factors like RRs [39,40] (Supplementary Figure 2a), to the entire subset of
123 RRs in the swapped cluster. We found high-confidence DNA binding targets for all of the NtrC-
124 like RRs (Supplementary Figure 2b, Supplementary Data 1,2). Aligning the high-confidence
125 targets from all of the NtrC-like RRs enabled us to manually assign a binding motif to the
126 homologs. Using the motifs identified from the high-confidence homologous hits, we were able to
127 query each genome for additional DNA binding targets (Supplementary Figure 2b, Figure 3a,c).
128 This analysis enabled us to hypothesize gene targets for the NtrC-like TCSs PP_1066, PP_0263,
129 PP_1401 to be upstream of PP_2453, PP_1188, and PP_1400 respectively. For all predicted
130 binding targets see Supplementary Table 2. As we were only able to find genomic targets for
131 AO356_22615 and AO356_25435 from *Pseudomonas fluorescens* FW300-N2C3
132 (Supplementary Figure 2b, Supplementary Data 1,2), we were unable to identify a binding motif
133 or a conserved genetic output for the NarL-like RRs by this method.
134 To determine the input signals for the cluster, we utilized a dataset of randomly barcoded
135 transposon libraries for next generation sequencing (RB Tn-seq) of *P. putida* KT2440 grown with
136 single carbon sources [41]. The genes with low fitness scores (fitness < -2) under a given growth
137 condition have condition-specific essentiality [42,43]. We observed that the gene clusters harboring
138 the RRs of interest each had low fitness scores for growth with either glutamic acid, succinic acid,
139 a-ketoglutaric acid, or butyric acid (Figure 3a). We speculated that genes and their corresponding
140 TCSs with synonymous fitness values represented a signaling cascade where the input signal
141 was the carbon source. Using this dataset, we proposed that the NtrC-like RRs, PP_1066,
142 PP_0263, and PP_1401 are regulated by glutamic acid [38], succinic acid, and a-ketoglutaric acid
143 [36,37] respectively. We also hypothesized that the NarL-like RR, PP_3551, is regulated by butyric
144 acid. As we were not able to identify binding sites or a DNA binding motif for PP_3551, we relied
145 on the gene cluster and fitness data to hypothesize an output for the butyric acid responsive TCS

146 [44,45]. The neighboring gene PP_3553 demonstrated a low fitness score consistent with the TCS,
147 making it a good candidate for the signaling cascade's output.

148 From the above results we were able to generate hypotheses for the signals and the
149 regulated genes of the TCSs in question. To validate these hypotheses, we generated GFP
150 reporter strains by tethering the upstream regions (~200bp) of the hypothesized output genes to
151 a GFP coding sequence on a broad host range plasmid. The *P. putida* KT2440 reporter strains
152 responded to the proposed signals with increased fluorescence above background (Figure 3b).
153 In concert with our predictions, when the reporter strains were stimulated in genetic backgrounds
154 absent of the corresponding transcriptional RR, fluorescence response was ablated (Figure 3b).
155 Interestingly, we observed that the WT strain bearing the p2453 reporter plasmid was slightly
156 activated by cultivation in butyric acid, indicating either sensing promiscuity by PP_1066's cognate
157 HK, PP_1067, or cross-talk with PP_3552, the cognate HK for PP_3551. We also observed that
158 when p3553 was grown with glutamic acid, fluorescence decreased below the minimal media (-)
159 levels, leading us to hypothesize that there is another component either within or external to the
160 identified TCSs that downregulated GFP expression.

161 Taken together, these results show that all of the input signals for the RRs in the swapped
162 cluster are short chain carboxylic acids (Figure 3c). We propose that these TCSs regulate either
163 uptake genes for carboxylic acids or metabolic genes involved in the TCA cycle or beta-oxidation.
164 We noted that the signal for HK-RR-promoter (PP_3552-PP_3551-pPP_3553) was unknown until
165 this study. We therefore rename the REC paralogs in the extra-familial subcluster the Carboxylic
166 Acid Responsive TCSs, CarSR I, II, III, IV (PP_0263, PP_1066, PP_1401, PP_3551).

167 *Exploring the sequence space of the RRs in the swapped cluster*

168 We next explored whether the CarSR TCSs were closely related using a coevolution
169 hypothesis, which dictates that if HK-RR pairs interact then the interacting residues should
170 coevolve across orthologs. In this view, closely related TCSs share more sequence space than

171 do TCSs that are distantly related [13,16]. We applied co-variance analysis [46,47] to HPt-CA-REC domains
172 of cognate HK-RR pairs to identify co-varying residues (Supplementary Figure 3) between HPt-
173 CA and REC domains. Our results are consistent with previous studies [10,15], where the first alpha
174 helix of the REC domain, previously shown to share points of contact with HPt domain [48], contained
175 the highest scoring residues. Importantly, the CarR cluster is well-insulated within the broader
176 NtrC-like, sigma54-dependent, cluster of the REC phylogenetic tree (Figure 2b,c), indicating that
177 its closest relatives are NtrC-like. We hypothesized that if the NarL-like and NtrC-like RRs in the
178 CarR subcluster could be engineered to switch specificity, despite their differences in TEDs and
179 sigma factor preference, then it was more likely that the RRs in the cluster share REC sequence
180 space.

181 To test if we could switch the specificity of the extra-familial CarR RRs, we applied site
182 directed mutagenesis at the positions of the predicted interacting sites to both PP_1066 (hereafter
183 referred to as RR_g) and PP_3551 (hereafter referred to as RR_b) (Figure 4a). Native or mutant RRs
184 (indicated by *) were expressed under the control of an arabinose inducible system (pBAD) and
185 GFP was driven by either the glutamic acid or butyric acid responsive promoters, hereafter
186 referred to as pRR_g (p2453) and pRR_b (p3553). To validate the use of the single plasmid system,
187 we complemented the ΔRR_g or ΔRR_b background strains with heterologous expression of RR_g or
188 RR_b. We determined that both of the heterologous RRs can respond to their expected signal in a
189 dose-dependent manner (Supplementary Figure 4). To ensure that there was no interference
190 between the two TCSs, the specificity-switch experiments were performed in double knockout
191 (dKO) background strains ($\Delta RR_g \Delta RR_b$). We observed leaky levels of GFP expression from native
192 RR_g-pRR_g in the dKO background in minimal media (-) (Figure 4b). In contrast, GFP expression
193 under the control of native RR_b-pRR_b in minimal media was tightly regulated (Figure 4b).
194 Consistent with our previous results (Figure 3b), we observed increased levels of GFP expression
195 when RR_g-pRR_g was grown with either glutamic acid and butyric acid (Figure 4b). We also
196 observed an increase in fluorescence when RR_b-pRR_b was grown with butyric acid, however we

197 did not observe the expected decrease in baseline fluorescence (Figure 3B) when RR_b-pRR_b was
198 grown with glutamic acid. When cultivated in minimal media with or without the addition of glutamic
199 acid, mutant RR_b*-pRR_b behaved like RR_g-pRR_g, demonstrating the characteristic leaky expression
200 in minimal media, and an increase in GFP expression in glutamic acid. However, when grown in
201 butyric acid, expression increased above the levels of growth in glutamic acid (Figure 4b). These
202 results suggest that RR_b* interacts with HK_g (PP_1067), but might also interact with HK_b (PP_3552)
203 when grown in butyric acid. GFP expression driven by RR_g*-pRR_g was leakier than its counterpart,
204 RR_b-pRR_b, which could be the result of either expression driven by a different promoter with
205 different dynamics or from interactions at a higher level of regulation. RR_g*-pRR_g demonstrated
206 the expected decrease in GFP expression below baseline observed in previous experiments
207 (Figure 3b) under glutamic acid growth conditions. Strikingly, RR_g*-pRR_g shows a statistically
208 significant increase in GFP expression when grown with butyric acid, suggesting that RR_g*
209 switched its specificity from HK_g to HK_b. Taken together, these results show, empirically, that
210 gradual changes in critical specificity determining residues made over time could be responsible
211 for the differentiation of the CarSR TCSs. Furthermore they demonstrate biologically relevant
212 evidence that these distinct extra-familial TCSs with different sigma factor dependencies, the
213 CarR RRs, are closely related.
214

215 *Interrogating an alternative hypothesis*

216 As an alternative to our working hypothesis, we conceived that if bacteria from other
217 distantly related phyla explored similar sequence space, the CarR subcluster could be derived
218 from HGT alone, and not a domain swap. To test this hypothesis, we queried the configuration of
219 RR families in bacterial phyla with high counts of NarL-like and NtrC-like RRs. Bacterial and
220 archaeal TCSs have been previously curated in a TCS census by Galperin et al. [19,49]. By querying
221 this dataset, we were able to find representative strains from Firmicutes, Bacteroides,

222 Acidobacteria, δ-Proteobacteria, and γ-Proteobacteria (*P. Putida* KT2440) that have high counts
223 of both NarL-like and NtrC-like RRs in their genomes (Supplementary Figure 5a). To test our
224 alternative hypothesis, we identified all RRs and their corresponding REC domains from the
225 representative species, made a REC domain-based phylogenetic tree, and mapped onto it the
226 corresponding RR family (Supplementary Figure 5b). We observed a similar configuration to
227 pseudomonads and proteobacteria trees (Figure 2b, Supplementary Figure 1), indicating that the
228 domain architecture we observe in present day bacteria is as ancient as the bacterial domain of
229 life, and the alternative hypothesis of HGT fails.
230

231 *Observation of HK domain architecture in related TCS*

232 On closer observation of the cognate HKs to RRs in the carboxylic acid sensing subcluster,
233 we found that the cognate HKs of the NtrC-like RRs are transmembrane bound with a d_Cache
234 domain and the cognate HKs to the NarL-like are cytosolic with a PAS domain. While both of
235 these sensing domains directly interact with small molecule ligands, they likely did not evolve
236 through lineage-specific expansion. To reconcile the discrepancy, we hypothesize that the HK
237 domain architecture observed in the pseudomonads carboxylic acid sensing subcluster is the
238 result of domain loss, where a common ancestor harbored both a periplasmic d_Cache domain
239 and a cytosolic PAS domain (Supplementary Figure 6). We anticipate that a closely related TCS
240 might exist in nature with the predicted domain architecture.

241 Discussion

242 In this work, we discover and provide evidence for a carboxylic acid sensing NtrC-like TCS
243 that underwent a domain swap, altering its sigma factor dependence (Figure 5). Our proposed
244 model assumes that an ancestral HK had promiscuous d_Cache and PAS carboxylic acid sensing
245 domains. Consistent with our characterization that the TCS senses carboxylic acids, both

246 d_Cache and PAS domains have been observed to interact directly with small molecules [50–52]. We
247 further propose that when the domain swap occured, the output gene was promiscuous within
248 carboxylic acid metabolism and could moonlight a beneficial function [53]. The *in vivo* mutational
249 screen provides evidence that insulated, extra-familial RRs within the same species share
250 sequence space and are closely related. In a previous study, an extra-familial specificity swap of
251 an Omp family HK enabled interaction with an Ntr family RR in *C. crescentus*, leading the authors
252 to hypothesize that a duplicated Ntr family TCS may have overlapped in sequence space with an
253 Omp family TCS [12]. Their results demonstrated how lineage-specific evolution of a duplicated TCS
254 facilitated insulation from unwanted cross-talk. In contrast, our work highlights how domain
255 swapping can alter the output TED (and classification) of an already well-insulated TCS.

256 NtrC-like RRs require sigma54 transcription factors to activate genes in their regulons,
257 whereas NarL-like RRs interact with sigma70. It is well documented that cellular levels of small
258 nucleotide, (p)ppGpp, in combination with the protein DksA heavily regulate RNA polymerase's
259 (RNAP) preference for sigma70 (housekeeping or extracytoplasmic function (ECF) family) [54–56] or
260 sigma54 [57–62]. While the exact environmental conditions needed for pseudomonads to reach the
261 appropriate levels of (p)ppGpp and DksA for sigma54 occupation of RNAP are unknown [61,63,64], it
262 can be reasoned that the switch is lifestyle dependent. We find the functional switch, resulting
263 from the identified domain swap, to be especially notable, because the TCSs implicated in this
264 study were found to be metabolite responsive systems, and could therefore play a major role in
265 survival and fitness. We speculate that the capability to consume carboxylic acids without the
266 regulatory constraints dictated by sigma factor dependence, may have given a fitness advantage
267 to the ancestral organism in which the swap originated.

268 Domain swapping in general is an important biological phenomenon that can impact all
269 domains of life [25]. Complicating our interpretation of bacterial evolution, HGT is known to mobilize
270 genes or entire gene clusters that can be subsequently integrated into foreign genomes [65–70]. The
271 narrative of HGT becomes even more entwined when we consider domain swapping. In this study

we applied a simple and straightforward strategy to identify domain swapping in a well-studied, modular signalling system. While we focused our attention on a single domain swap, our results (Figure 2, Supplementary Figures 1, 6) and results of previous studies [7,24] suggest that domain swapping in TCS is prevalent. Our strategy, however, is not limited to signalling systems. Given the potential ubiquity of domain swapping, we propose that this strategy can be applied to other complex modular proteins and systems to detangle their intricate evolutions and cryptic functions.

Nature's pre-engineered enzymes and proteins provide attractive parts for practical applications [71,72] and potential answers to plug-and-play biology via synthetic domain swapping [73,74]. One key and critical component of synthetic domain swapping is identifying domain boundaries for a seamless swap [74–76]. In a previous study a library based approach was implemented to identify domain boundaries between REC domains and TEDs for *E. coli* RRs, however their successes were only shown for inter-family TED swaps [76]. In this study, we demonstrate that nature devised the appropriate domain boundary between a REC domain and an extra-familial TED to engineer a functional signaling cascade. Such evolutionarily driven domain swaps could be an untapped source for identifying optimal domain boundaries in synthetic biology applications.

Taken together, this work establishes and validates domain swapping as a mechanism for neofunctionalization of modular genes. By applying hypotheses grounded in the established theories of evolution of TSCs, we were able to link an orphaned sigma70-dependent TCS to a well-described family of sigma54-dependent TCSs. Our results validate the close relationship between systems that would have otherwise been described as distantly related. Ultimately, our work provides evidence for an extension to the current model of neofunctionalization to include functional part sharing via domain swapping.

Acknowledgements

We would like to thank the following individuals for their contribution to this body of work for either helping with lab tasks, Andrew Lau, Julie Lake, Rodrigo Frogeso, and Joyce Luk; providing

valuable insight through discussion about evolution, fitness, and bacterial transcriptional regulation, Ankita Kothari, Pablo Cruz-Morales, Mitchell G. Thomson; providing expertise toward the development of high-throughput cloning, protein purification, and NGS techniques, Nurgul Kaplan, Jennifer Chiniquy, Joel M. Guenther, and Brett Garabedian; or providing us with capabilities to perform high throughput fitness assays, Mitchell G. Thompson, Morgan Price, Adam Deutschbauer.

Funding source:

This work was part of the ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Science Focus Area Program at Lawrence Berkeley National Laboratory (LBNL) and is supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231 between LBNL and the U. S. Department of Energy. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

Author Contributions

M.E.G., A.K., and H.Z. performed phylogenetic computational analyses; M.E.G. and L.R. designed and built expression strains for DAP-seq; M.E.G. developed and conducted the automated pipeline for DAP-seq experiments; A.K. analyzed DAP-seq data; M.E.G. ran co-variance analysis; M.E.G. and V.F. conducted and analyzed *in vivo GFP reporter* experiments;

M.E.G. designed the experiments and wrote the first draft. M.R.I. performed single carbon source fitness assays. J.D.K. provided support and resources. A.M. provided resources supervision and support. All authors reviewed and edited the final draft.

Declarations of Interest

J.D.K. has financial interests in Amyris, Lygos, Demetrix, Napigen and Maple Bio.

Figure and Figure Legends

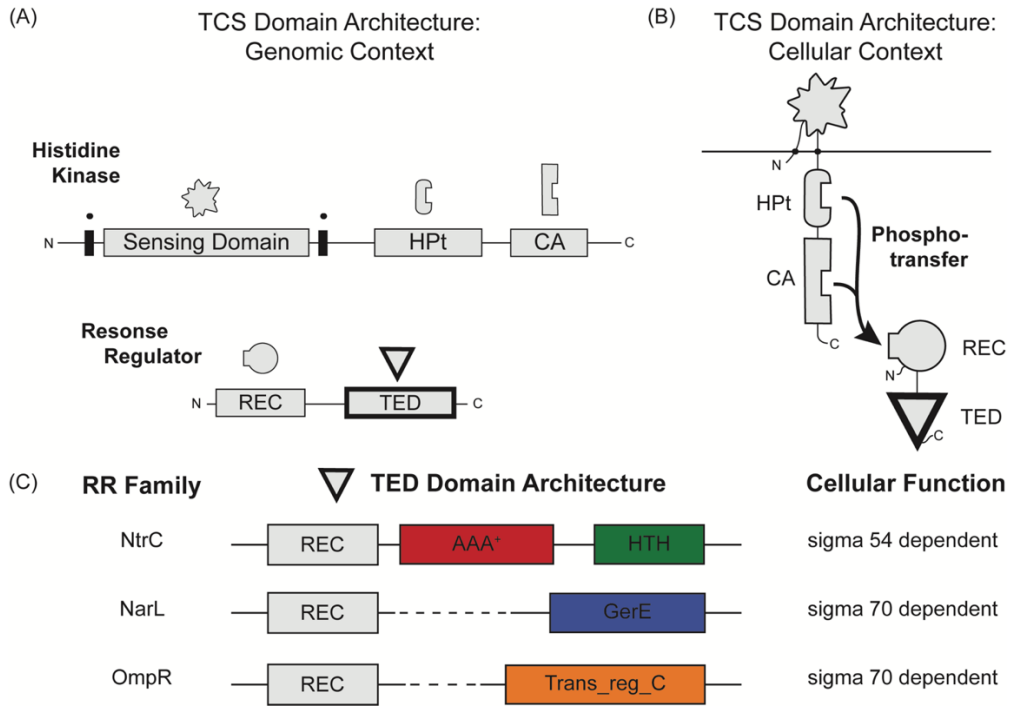


Figure 1: ***Domain architecture of TCSs with transcriptional effector domains (TEDs):*** Modularity of canonical TCS in its genomic (A) and cellular (B) contexts. (C) RR families designated by the domain architecture of transcriptional effector domains (TEDs) found in TCS relevant to this study annotated by their cellular function.

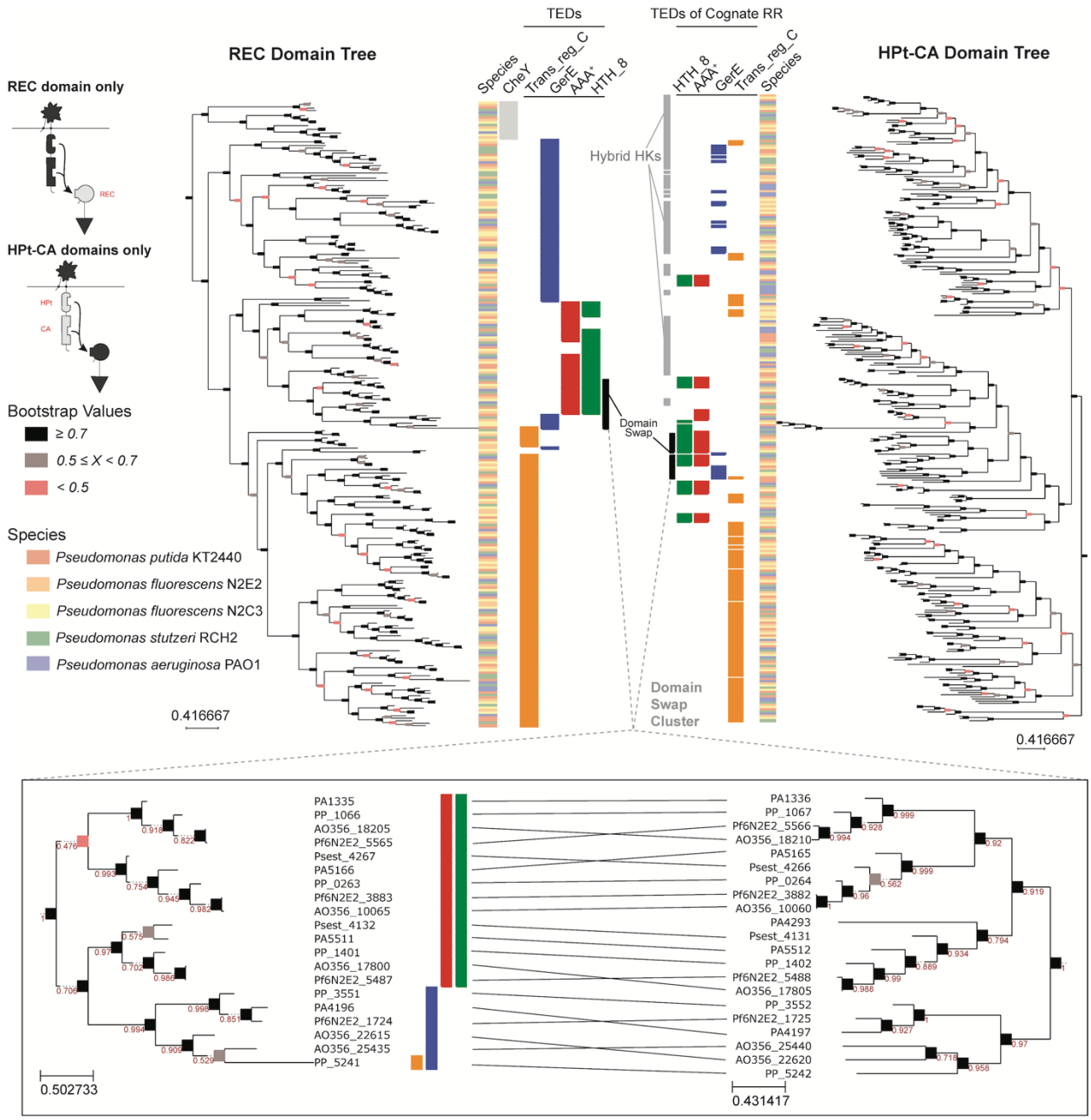
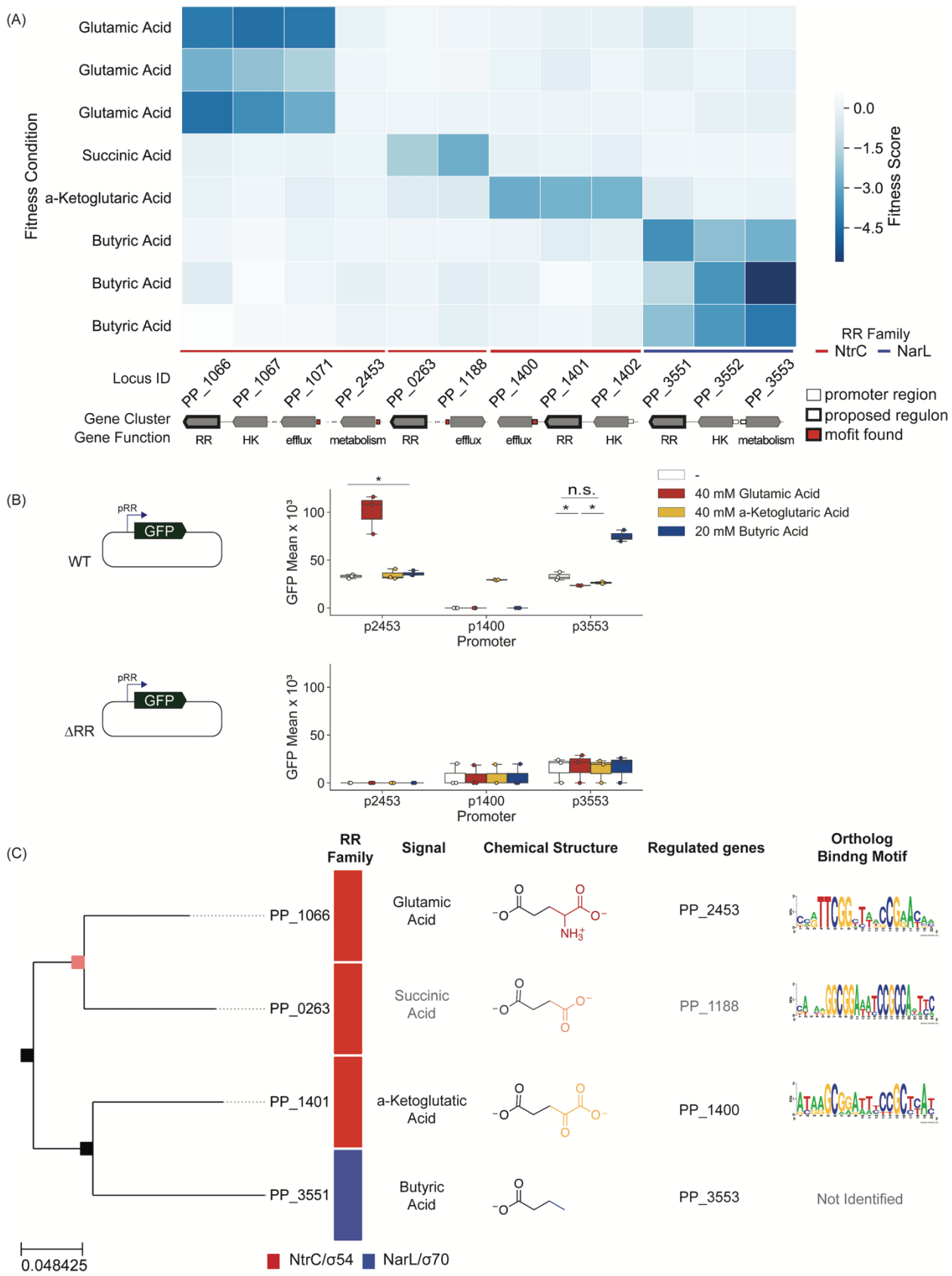


Figure 2: **Phylogenetic analysis of TCS domains from *Pseudomonas* RRs reveals TED domain swap:** Phylogenetic trees of REC (left) and HPt-CA (right) domains of RRs and HKs from select Pseudomonads. Trees are annotated with species and transcriptional effector domain (TED) at the leaves. Hybrid HKs on the HPt-CA tree are highlighted in grey, and are notably associated with TCS that have GerE TEDs. The domain swap at the focus of this work is highlighted in black. A close-up of the domain swap cluster for REC (left) and HPt-CA (right)

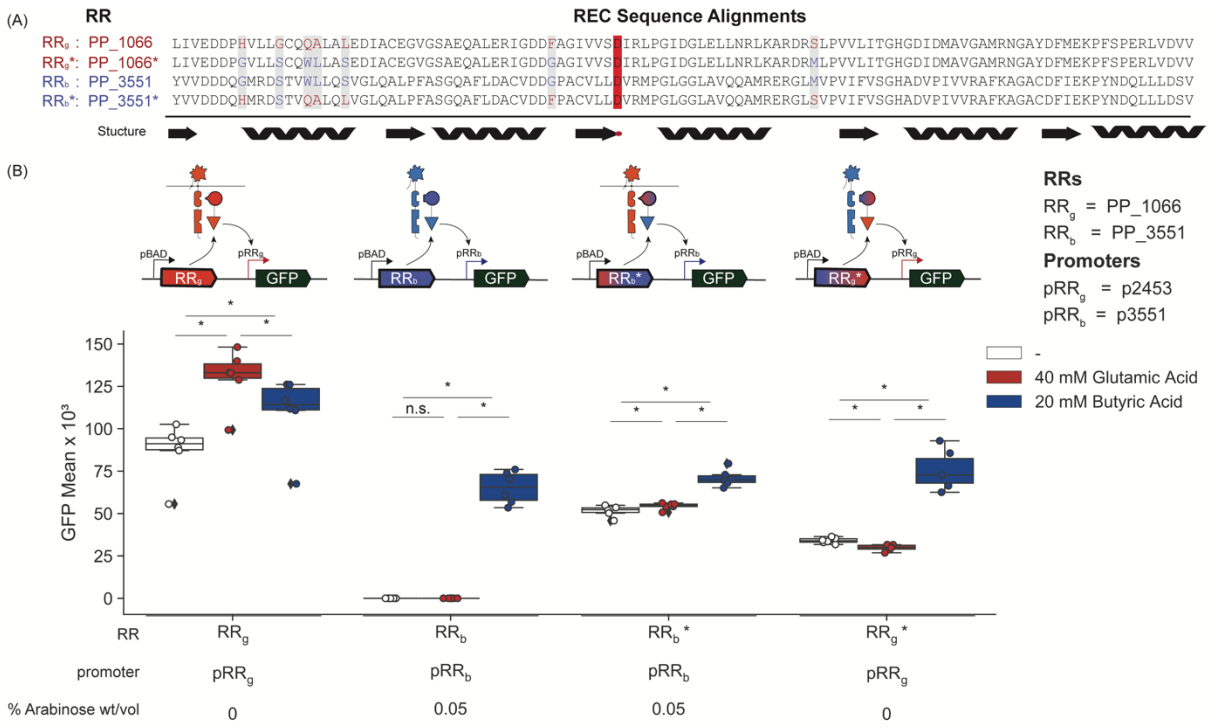
339 domains trees annotated with protein name and/or TED. Lines are drawn to match predicted

340 cognate HK-RR pairs.


(A)
Fitness Condition
Glutamic Acid
Glutamic Acid
Glutamic Acid
Succinic Acid
a-Ketoglutaric Acid
Butyric Acid
Butyric Acid
Butyric Acid
Fitness Score
0.0
−1.5
−3.0
−4.5
RR Family
NtrC
NarL
Locus ID
PP_1066
PP_1067
PP_1071
PP_2453
PP_0263
PP_1188
PP_1400
PP_1401
PP_1402
PP_3551
PP_3552
PP_3553
Gene Cluster
Gene Function
RR
HK
efflux
metabolism
RR
efflux
efflux
RR
HK
RR
HK
metabolism
promoter region
proposed regulon
mofit found
(B)
pRR
GFP
WT
ΔRR
GFP Mean x 10³
100
50
0
p2453
p1400
p3553
Promoter
*
n.s.
-
40 mM Glutamic Acid
40 mM a-Ketoglutaric Acid
20 mM Butyric Acid
(C)
RR Family
Signal
Chemical Structure
Regulated genes
Ortholog Bindng Motif
PP_1066
PP_0263
PP_1401
PP_3551
Glutamic Acid
Succinic Acid
a-Ketoglutatic Acid
Butyric Acid
PP_2453
PP_1188
PP_1400
PP_3553
Not Identified
0.048425
NtrC/σ54
NarL/σ70

342 Figure 3: ***Determining the functional similarity of the RRs in the swapped cluster***: (A)
343 Heatmap of fitness data for TCS gene clusters within the domain swapped cluster. A low fitness
344 score (fitness< - 2) indicates a gene has condition-specific essentiality. DAP-seq results are
345 briefly summarized with a red box upstream of the gene in which an orthogonal DNA binding motif
346 was identified. (B) GFP reporter plasmids in WT and ΔRR backgrounds were grown in minimal
347 media with glucose plus a second carbon source - glutamic acid, a-ketoglutaric acid or butyric
348 acid. Fluorescence was measured by flow cytometry and measurements are reported as GFP
349 Mean x 10^3after gating. Statistical significance was determined between conditions by t-test (* =
350 p-value < 0.05, n.s. = not significant). (C) Summary of findings from functional genomics studies.
351 The relationship of the *P. putida* TCSs is shown by a pruned phylogenetic tree to the left, and
352 results, RR family, validated sigal, signal chemical structure, validated regulated genes,
353 orthogonal binding motif identified by DAP-seq are drawn as a table to the right. If the findings
354 were not validated by reporter assay, the results are displayed in grey.
355

356



357

358 Figure 4: ***Exploring the sequence space of the RRs in the swapped cluster:*** (A) Sequence
359 alignments of native RRs PP_1066 (RR_g), PP_3551 (RR_b), and mutated RRs PP_1066* (RR_g*),
360 PP_3551* (RR_b*). Residues with high co-variance scores > 1.1 are marked in a grey background
361 and red text for RR_g-derived residues or blue text for RR_b-derived residues. Structural information
362 for the REC domain is shown below the sequences, where arrows are β-barrels and ribbons are
363 α-helices. (B) RR_g, driving glutamic acid responsive promoter (pRR_g) switches its specificity to
364 butyric acid responsive when mutated to RR_g*. RR_b, driving butyric acid responsive promoter
365 (pRR_b) partially switches its specificity to glutamic acid responsive when mutated to RR_b* .
366 Statistical significance was determined between conditions by t-test (* = p-value < 0.05, n.s. = not
367 significant).

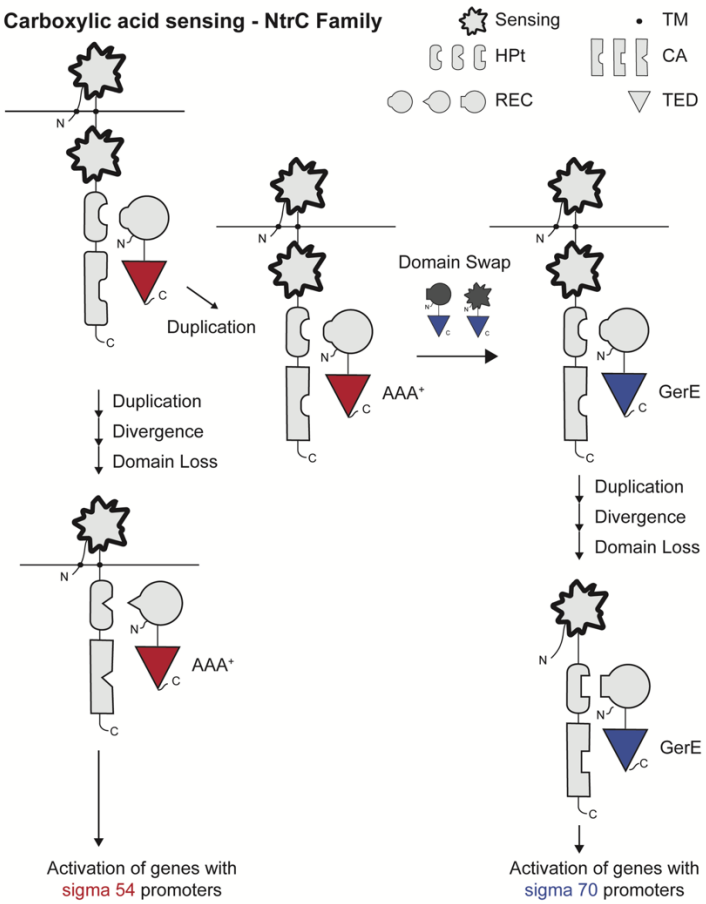


Figure 5: ***Proposed route for evolution of carboxylic acid sensing TCS with alternative sigma factor dependence*:** Within an ancestral host, an Ntr-like TCS with carboxylic acid sensing capabilities was duplicated. A transcription factor with a GerE DNA binding domain recombined with the duplicated gene to form an active and functional Ntr-like TCS with a GerE DNA binding domain. Both TCSs encoded in the same genome undergo duplication, divergence, and domain loss to achieve insulated pathways with insulated signal detection. Each system in its modern form can detect unique carboxylic acids and activate unique sets of genes. The key difference between the related systems is they interact with different types of promoters either sigma54 or sigma70, which are known to be active under different environmental constraints. We hypothesize that the ancestral strain in which the domain swap originated might have benefited from a lack of environmental constraints against consuming available environmental resources.

380 # Methods

381 **Phylogenetic analysis of REC and HPt-CA trees**

382 RRs and HKs were identified with hmmsearch from HMMER v3.1b2 [77] to search for all proteins
383 with Response_reg (PF00072) and HisKA (PF00512) domains. Alternatively, HKs and RRs were
384 identified with the Microbial Signal Transduction database (MISTDB) [78]. Using HMMER again, we
385 queried the domain architecture of the signaling proteins and selected for RRs with HTH_8,
386 Trans_reg_C, AAA+, GerE, HTH_18 (AraC) domains. REC and HPt-CA domain sequences were
387 extracted from whole sequences based on coordinates determined by the hmmsearch. Domains
388 were aligned using the MAFFT-LINSI algorithm from MAFFT v7.310 [79]. Phylogenetic trees were
389 constructed using FastTree 2 [80]. Trees were visualized and annotated using python based ETE3
390 [81].

391 **Automated DNA affinity purification - seq**

392 *DNA preparation for NGS*

393 Pseudomonas isolates were cultured in either LB or minimal media (see Supplementary Table 3
394 for strain specific minimal media recipes). Genomic DNA was purified with a promega wizard
395 genomic preparation kit (Promega, Madison, WI). DNA was sheared with covaris miniTUBE
396 (Covaris, Woburn, MA) to an average size of 200 bp. The DNA quality was confirmed by
397 Bioanalyzer high sensitivity DNA kit (Agilent, Santa Clara). Sheared DNA was then adapter-
398 ligated (AL) with NEBnext Ultra ii Library Preparation kit (New England Biolabs, Ipswich, MA). AL-
399 DNA quality was again confirmed by Bioanalyzer high sensitivity DNA kit (Agilent, Santa Clara).
400 AL-DNA was stored at -20˚C until required for downstream use.
401

402 *Expression Strain Design*

403 Pet28 expression vectors with N-terminal 6x-His-tagged RRs were cloned by Gibson assembly [82].
404 Plasmid design was facilitated by j5 DNA assembly design [83] (diva.jbei.org), see Supplementary
405 Table 4 for primers.

406

407 *Automated DNA affinity purification*

408 Quadruplicates of expression strains were grown in autoinduction media (Zyp-5052 [84]) at 37˚C,
409 250 RPM, for 5-6 hours and then transferred to grow at 17˚C, 250 RPM, overnight. Cell pellets
410 were harvested and lysed at 37˚C for 1 hour in a lysis buffer (1X TBS, 100 μM PMSF (Millipore
411 Sigma, Burlington MA), 2.5 units/mL Benzonase nuclease (Millipore Sigma, Burlington MA), 1
412 mg/mL Lysozyme (Millipore Sigma, Burlington MA)). Lysed cells were then clarified by
413 centrifugation at 3214 x g and further filtered in 96-well filter plates by centrifugation at 1800 x g.
414 To enable high-throughput processing, protein-DNA purification steps were performed with IMAC
415 resin pipette tips (PhyNexus, San Jose, CA) using a custom automated platform with the Biomek
416 FX liquid handler (Beckman Coulter, Indianapolis, IN). The expressed RRs were individually
417 bound to metal affinity resin embedded within the IMAC resin pipette tips and washed in a wash
418 buffer (1X TBS, 10 mM Imidazole, 0.1% Tween 20). The bead bound RRs were then mixed with
419 60μL of DNA binding buffer (1X TBS, 10 mM magnesium chloride, 0.4 ng/μL AL-DNA, with or
420 without 50 mM acetyl phosphate (split into duplicates)). The protein bound to its target DNA was
421 then enriched in an enrichment buffer (1X TBS, 10 mM Imidazole, 0.1% Tween 20) and eluted in
422 an elution buffer (1X TBS, 180 mM Imidazole). The elution was stored at -20˚C for a minimum of
423 one day and up to a week before proceeding to the NGS library generation. See supplementary
424 methods for detailed protocol.

425

426 *NGS Library Generation*

427 3.2 µL of the elution from the previous step was added to 3.5 µL SYBR green ssoAdvanced
428 (Biorad, Hercules, CA) and 0.15 µL of each dual indexed NGS primers. NGS libraries were
429 prepared by following the protocols for fluorescent amplification of NGS libraries [85]. Pooled libraries
430 were sequenced by Illumina NovaSeq 6000 SP (100 cycles) (Illumina, San Diego, CA).

431

432 *DAP-seq data analysis*

433 Sequenced reads were processed by a computational DAP-seq analysis pipeline as follows.
434 Adapters and low-quality bases were trimmed and reads shorter than 30 bp were filtered out using
435 Trimmomatic v.0.36 [86]. The resulting reads were checked for contamination using FOCUS [87]. Then
436 the reads were aligned to the corresponding *Pseudomonas* spp. genome using Bowtie v1.1.2 [88]
437 with –m 1 parameter (report reads with single alignment only). Resulting SAM files were converted
438 to BAM format and sorted using samtools v 0.1.19 [89]. Peak calling was performed using SPP
439 1.16.0 [90] with false discovery rate threshold of 0.01 and MLE enrichment ratio threshold of 4.0.
440 Enriched motifs were discovered in genome fragments corresponding to the peaks using MEME
441 [91] with parameters –mod anr –minw 12 –maxw 30 –revcomp –pal –nmotifs 1. Source code of the
442 DAP-seq analysis pipeline is available at https://github.com/novichkov-lab/dap-seq-utils.

443 For conserved RRs with small numbers of high-confidence peaks (1-2 per genome), binding
444 motifs were predicted manually by comparative genomics approach. Orthologous RRs were
445 identified by OrthoFinder2 [92]. For each of orthologous RRs, one genome fragment corresponding
446 to the peak with the highest enrichment value was selected for motif search. Conserved motifs
447 were discovered using the SignalX tool from GenomeExplorer package [93] with the “inverted repeat”
448 option.

449

450 **Fitness Experiments**

451 The *Pseudomonas putida* KT2440 RB-TnSeq library, JBEI-1 [94], was thawed, inoculated into 25
452 mL of LB medium supplemented with kanamycin, and grown to OD_{600} of 0.5. Three 1 mL
453 samples were taken after this step to serve as t_0 records of barcode abundance. The library was
454 then washed via centrifugation and resuspension in an equal volume of MOPS minimal medium.
455 The washed cells were then diluted 1:50 in MOPS minimal medium with 10 mM L-glutamate
456 serving as the sole carbon source. The library was cultured in a 96-well deep well plate sealed
457 with a gas permeable membrane (VWR, USA). The plate was shaken (700 rpm) in an INFORS
458 HT Multitron (Infors USA Inc.) at 30 °C for 24 hrs. Duplicate 600 µL samples were then
459 combined and BarSeq analysis was conducted as described previously [42,95,96].
460 Single carbon source fitness data is available at http://fit.genomics.lbl.gov [41].

461 **Co-variance analysis**

462 Cognate RRs and HKs from *Pseudomonas* and *E. coli* strains were identified as pairs if they were
463 found neighboring each other in their respective genomes. HPt (HisKA), CA (HATPase_C), and
464 REC (Response_reg) domain boundaries were determined with hmmsearch from HMMER
465 v3.1b2 [77]. Fasta files of concatenated HPt-CA-REC domains from cognate and randomized HK-
466 RR pairs were aligned with the MAFFT-LINSI algorithm from MAFFT v7.310 [79]. Alignment files
467 were then queried for coevolution with the ProDy *Evol* suite [46,47] in python and were plotted in a
468 heatmap. The highest scoring residues > 1.1 were used to inform hypotheses for specificity switch
469 strains.

470 **GFP reporter strain generation and assays**

471 *Knockout strain generation*

472 1000bp homology fragments upstream and downstream of the target gene were cloned into
473 plasmid pKS18. Plasmids were then transformed into E. coli S17 and then mated into P. putida

474 via conjugation. Transconjugants were selected for on LB agar plates supplemented with 30
475 mg/ml kanamycin and 30 mg/ml chloramphenicol. Transconjugants were then grown overnight
476 on LB media and were then plated on LB agar with no NaCl that was supplemented with 10%
477 (wt/vol) sucrose. Putative deletions were screened on LB agar with no NaCl supplemented with
478 10% (wt/vol) sucrose and LB agar plate with kanamycin. Colonies that grew in the presence of
479 sucrose but had no resistance to kanamycin were further tested via PCR with primers flanking
480 the target gene to confirm gene deletion.

481 *GFP reporter strains*

482 Promoter boundaries for p2453, p1400, p3553 were identified as the region just upstream of the
483 gene's start codon up until the start or stop codon of the next nearest gene. The promoters were
484 cloned upstream of GFP on a broad host range plasmid with BBR1 origin and Kanamycin
485 resistance with Gibson cloning [82], primers in Supplementary Table 4. The plasmids were
486 transformed into *P. putida* KT2440 or *P. putida* KT2440 mutant strains by electroporation. Three
487 biological replicates of each strain were cultured in LB and stored with 25% (vol/vol) glycerol at -
488 80 ˚C. Complementation plasmids (GFP reporter plasmids with full length RR driven by pBAD
489 promoter and constitutively expressed AraC) were combinatorially built leveraging Golden Gate
490 cloning [97] and j5 DNA assembly design [83] (diva.jbei.org), primers in Supplementary Table 4. The
491 plasmids were transformed into knockout strains of *P. putida* KT2440 by electroporation. 3-6
492 biological replicates of each strain were cultured in LB and stored with 25 % (v/v) glycerol at -80
493 ˚C.

494 *RRs with switched specificity*

495 Gene blocks (TWIST Biosciences, San Francisco, CA) of REC domains (Supplementary Table
496 3) with co-varying mutations (co-variation score > 1.1) were cloned into the complementation
497 plasmids with Gibson assembly [82]. The plasmids were transformed into *P. putida* KT2440 or

498 knockout strains of *P. putida* KT2440 by electroporation. Six biological replicates of each strain
499 were cultured in LB and stored with 25 % (vol/vol) glycerol at -80 ˚C.

500 *GFP reporter Assays*

501 Reporter strains were adapted to M9 minimal media (MM) (see supplemental methods for strain
502 specific minimal media recipes) supplemented with 0.5 % (wt/vol) glucose as the sole carbon
503 source in 3 overnight passages, and were stored in MM at -80˚C in 25 % (vol/vol) glycerol.
504 Adapted strains were cultured in MM + 0.5 % (wt/vol) glucose and passaged to MM + 0.5 %
505 (wt/vol) glucose with or without a second carbon-source (40 mM glutamic Acid, 40 mM a-
506 ketoglutaric acid, or 20 mM butyric acid, unless otherwise specified). After 24-hours of growth,
507 fluorescence was measured by flow cytometry on the BD Accuri C6 (BD Biosciences, San Jose,
508 CA). Autofluorescence was gated out with FlowJo (BD Biosciences, San Jose, CA), using a non-
509 fluorescent strain of P. putida KT2440 carrying an empty vector plasmid for reference. To remove
510 noise, the GFP mean for samples with less than 150 events after gating was set to 0. Otherwise,
511 the GFP mean of the remaining events after gating was reported. Statistical significance was
512 determined by T-test between replicates.

513 **Resource Availability**

514 *Data and Code Availability*

515 Source code of the DAP-seq analysis pipeline is available at https://github.com/novichkov-
516 lab/dap-seq-utils

517 DAP-seq data is have been deposited in NCBI's Gene Expression Omnibus [98] and are accessible
518 through GEO Series accession number GSE157075
519 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157075)

520 Single carbon source fitness data is available at http://fit.genomics.lbl.gov

521 # References

522 1. Sankhe, G. D., Dixit, N. M. & Saini, D. K. Activation of Bacterial Histidine Kinases: Insights into
523 the Kinetics of the cis Autophosphorylation Mechanism. *mSphere* **3,** (2018).

524 2. Zschiedrich, C. P., Keidel, V. & Szurmant, H. Molecular Mechanisms of Two-Component Signal
525 Transduction. *J. Mol. Biol.* **428,** 3752–3775 (2016).

526 3. Galperin, M. Y. A census of membrane-bound and intracellular signal transduction proteins in
527 bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol.* **5,** 35 (2005).

528 4. Wuichet, K., Cantwell, B. J. & Zhulin, I. B. Evolution and phyletic distribution of two-component
529 signal transduction systems. *Curr. Opin. Microbiol.* **13,** 219–225 (2010).

530 5. Galperin, M. Y., Higdon, R. & Kolker, E. Interplay of heritage and habitat in the distribution of
531 bacterial signal transduction systems. *Mol. Biosyst.* **6,** 721–728 (2010).

532 6. Jung, K., Fried, L., Behr, S. & Heermann, R. Histidine kinases and response regulators in
533 networks. *Curr. Opin. Microbiol.* **15,** 118–124 (2012).

534 7. Grebe, T. W. & Stock, J. B. The histidine protein kinase superfamily. *Adv. Microb. Physiol.* **41,**
535 139–227 (1999).

536 8. Podgornaia, A. I. & Laub, M. T. Determinants of specificity in two-component signal
537 transduction. *Curr. Opin. Microbiol.* **16,** 156–162 (2013).

538 9. Choi, K. & Kim, S. Building interacting partner predictors using co-varying residue pairs
539 between histidine kinase and response regulator pairs of 48 bacterial two-component systems.
540 *Proteins* **79,** 1118–1131 (2011).

541 10. Laub, M. T., Biondi, E. G. & Skerker, J. M. Phosphotransfer profiling: systematic mapping of
542 two-component signal transduction pathways and phosphorelays. *Meth. Enzymol.* **423,** 531–548
543 (2007).

544 11. Salazar, M. E. & Laub, M. T. Temporal and evolutionary dynamics of two-component signaling
545 pathways. *Curr. Opin. Microbiol.* **24,** 7–14 (2015).

546 12. Capra, E. J., Perchuk, B. S., Skerker, J. M. & Laub, M. T. Adaptive mutations that prevent
547 crosstalk enable the expansion of paralogous signaling protein families. *Cell* **150,** 222–232
548 (2012).

549 13. Capra, E. J. & Laub, M. T. Evolution of two-component signal transduction systems. *Annu.*
550 *Rev. Microbiol.* **66,** 325–347 (2012).

551 14. Laub, M. T. & Goulian, M. Specificity in two-component signal transduction pathways. *Annu.*
552 *Rev. Genet.* **41,** 121–145 (2007).

553 15. Skerker, J. M. *et al.* Rewiring the specificity of two-component signal transduction systems.
554 *Cell* **133,** 1043–1054 (2008).

555 16. McClune, C. J. & Laub, M. T. Constraints on the expansion of paralogous protein families.
556 *Curr. Biol.* **30,** R460–R464 (2020).

557 17. Herrou, J., Crosson, S. & Fiebig, A. Structure and function of HWE/HisKA2-family sensor
558 histidine kinases. *Curr. Opin. Microbiol.* **36,** 47–54 (2017).

559 18. Lori, C., Kaczmarczyk, A., de Jong, I. & Jenal, U. A Single-Domain Response Regulator
560 Functions as an Integrating Hub To Coordinate General Stress Response and Development in
561 Alphaproteobacteria. *MBio* **9,** (2018).

562 19. Galperin, M. Y. Diversity of structure and function of response regulator output domains. *Curr.*
563 *Opin. Microbiol.* **13,** 150–159 (2010).

564 20. Padilla-Vaca, F., Mondragón-Jaimes, V. & Franco, B. General Aspects of Two-Component
565 Regulatory Circuits in Bacteria: Domains, Signals and Roles. *Curr. Protein Pept. Sci.* **18,** 990–
566 1004 (2017).

567 21. Dutta, R., Qin, L. & Inouye, M. Histidine kinases: diversity of domain organization. *Mol.*
568 *Microbiol.* **34,** 633–640 (1999).

569 22. Kim, S., Hirakawa, H., Muta, S. & Kuhara, S. Identification and classification of a two-
570 component system based on domain structures in bacteria and differences in domain structure
571 between Gram-positive and Gram-negative bacteria. *Biosci. Biotechnol. Biochem.* **74,** 716–720
572 (2010).

573 23. Ortega, Á., Zhulin, I. B. & Krell, T. Sensory repertoire of bacterial chemoreceptors. *Microbiol.*
574 *Mol. Biol. Rev.* **81,** (2017).

575 24. Alm, E., Huang, K. & Arkin, A. The evolution of two-component systems in bacteria reveals
576 different strategies for niche adaptation. *PLoS Comput. Biol.* **2,** e143 (2006).

577 25. Forslund, S. K., Kaduk, M. & Sonnhammer, E. L. L. Evolution of protein domain architectures.
578 *Methods Mol. Biol.* **1910,** 469–504 (2019).

579 26. Lai, R.-Z. & Parkinson, J. S. Monitoring Two-Component Sensor Kinases with a Chemotaxis
580 Signal Readout. *Methods Mol. Biol.* **1729,** 127–135 (2018).

581 27. Briegel, A. *et al.* Universal architecture of bacterial chemoreceptor arrays. *Proc Natl Acad Sci*
582 *USA* **106,** 17181–17186 (2009).

583 28. Ryjenkov, D. A., Tarutina, M., Moskvin, O. V. & Gomelsky, M. Cyclic diguanylate is a
584 ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein
585 domain. *J. Bacteriol.* **187,** 1792–1798 (2005).

586 29. Chen, Y.-T., Chang, H. Y., Lu, C. L. & Peng, H.-L. Evolutionary analysis of the two-component
587 systems in Pseudomonas aeruginosa PAO1. *J. Mol. Evol.* **59,** 725–737 (2004).

588 30. Voordeckers, K., Pougach, K. & Verstrepen, K. J. How do regulatory networks evolve and
589 expand throughout evolution? *Curr. Opin. Biotechnol.* **34,** 180–188 (2015).

590 31. Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1,** 16048 (2016).

591 32. Nishijyo, T., Haas, D. & Itoh, Y. The CbrA-CbrB two-component regulatory system controls
592 the utilization of multiple carbon and nitrogen sources in Pseudomonas aeruginosa. *Mol.*
593 *Microbiol.* **40,** 917–931 (2001).

594 33. Cases, I., Ussery, D. W. & De Lorenzo, V. The σ^{54} regulon (sigmulon) of*Pseudomonas putida*.
595 *Environ. Microbiol.* **5,** 1281–1293 (2003).

596 34. Leech, A. J., Sprinkle, A., Wood, L., Wozniak, D. J. & Ohman, D. E. The NtrC family regulator
597 AlgB, which controls alginate biosynthesis in mucoid Pseudomonas aeruginosa, binds directly to
598 the algD promoter. *J. Bacteriol.* **190,** 581–589 (2008).

599 35. Hervás, A. B., Canosa, I., Little, R., Dixon, R. & Santero, E. NtrC-dependent regulatory
600 network for nitrogen assimilation in *Pseudomonas putida*. *J. Bacteriol.* **191,** 6123–6135 (2009).

601 36. Tatke, G., Kumari, H., Silva-Herzog, E., Ramirez, L. & Mathee, K. Pseudomonas aeruginosa
602 MifS-MifR Two-Component System Is Specific for α-Ketoglutarate Utilization. *PLoS ONE* **10,**
603 e0129629 (2015).

604 37. Lundgren, B. R. *et al.* Genetic analysis of the assimilation of C5-dicarboxylic acids in
605 Pseudomonas aeruginosa PAO1. *J. Bacteriol.* **196,** 2543–2551 (2014).

606 38. Sonawane, A. M., Singh, B. & Röhm, K.-H. The AauR-AauS two-component system regulates
607 uptake and metabolism of acidic amino acids in Pseudomonas putida. *Appl. Environ. Microbiol.*
608 **72,** 6569–6577 (2006).

609 39. Rajeev, L., Garber, M. E. & Mukhopadhyay, A. Tools to map target genes of bacterial two-
610 component system response regulators. *Environ. Microbiol. Rep.* **12,** 267–276 (2020).

611 40. Garber, M. E. *et al.* Multiple signaling systems target a core set of transition metal homeostasis
612 genes using similar binding motifs. *Mol. Microbiol.* **107,** 704–717 (2018).

613 41. Thompson, M. G. *et al.* Functional analysis of the fatty acid and alcohol metabolism of
614 Pseudomonas putida using RB-TnSeq. *Appl. Environ. Microbiol.* (2020).
615 doi:10.1128/AEM.01665-20

616 42. Wetmore, K. M. *et al.* Rapid quantification of mutant fitness in diverse bacteria by sequencing
617 randomly bar-coded transposons. *MBio* **6,** e00306-15 (2015).

618 43. Price, M. N. *et al.* Mutant phenotypes for thousands of bacterial genes of unknown function.
619 *Nature* **557,** 503–509 (2018).

620 44. Thompson, M. G. *et al.* Robust characterization of two distinct glutarate sensing transcription
621 factors of *Pseudomonas putida* L-lysine metabolism. *ACS Synth. Biol.* (2019).
622 doi:10.1021/acssynbio.9b00255

623 45. Thompson, M. G. *et al.* Identification, Characterization, and Application of a Highly Sensitive
624 Lactam Biosensor from Pseudomonas putida. *ACS Synth. Biol.* **9,** 53–62 (2020).

625 46. Bakan, A., Meireles, L. M. & Bahar, I. ProDy: protein dynamics inferred from theory and
626 experiments. *Bioinformatics* **27,** 1575–1577 (2011).

627 47. Bakan, A. *et al.* Evol and ProDy for bridging protein sequence evolution and structural
628 dynamics. *Bioinformatics* **30,** 2681–2683 (2014).

629 48. Jacob-Dubuisson, F., Mechaly, A., Betton, J.-M. & Antoine, R. Structural insights into the
630 signalling mechanisms of two-component systems. *Nat. Rev. Microbiol.* **16,** 585–593 (2018).

631 49. Galperin, M. Y. Structural classification of bacterial response regulators: diversity of output
632 domains and domain combinations. *J. Bacteriol.* **188,** 4169–4182 (2006).

633 50. Henry, J. T. & Crosson, S. Ligand-binding PAS domains in a genomic, cellular, and structural
634 context. *Annu. Rev. Microbiol.* **65,** 261–286 (2011).

635 51. Brewster, J. L. *et al.* Structural basis for ligand recognition by a Cache chemosensory domain
636 that mediates carboxylate sensing in Pseudomonas syringae. *Sci. Rep.* **6,** 35198 (2016).

637 52. Gavira, J. A. *et al.* How bacterial chemoreceptors evolve novel ligand specificities. *MBio* **11,**
638 (2020).

639 53. Pougach, K. *et al.* Duplication of a promiscuous transcription factor drives the emergence of
640 a new regulatory network. *Nat. Commun.* **5,** 4868 (2014).

641 54. Casas-Pastor, D. *et al.* Expansion and re-classification of the extracytoplasmic function (ECF)
642 σ factor family. *BioRxiv* (2019). doi:10.1101/2019.12.11.873521

643 55. Lonetto, M. A. *et al.* Identification of a contact site for different transcription activators in region
644 4 of the Escherichia coli RNA polymerase sigma70 subunit. *J. Mol. Biol.* **284,** 1353–1365 (1998).

645 56. Lee, D. J., Minchin, S. D. & Busby, S. J. W. Activating transcription in bacteria. *Annu. Rev.*
646 *Microbiol.* **66,** 125–152 (2012).

647 57. Bernardo, L. M. D., Johansson, L. U. M., Skärfstad, E. & Shingler, V. sigma54-promoter
648 discrimination and regulation by ppGpp and DksA. *J. Biol. Chem.* **284,** 828–838 (2009).

649 58. Dalebroux, Z. D. & Swanson, M. S. ppGpp: magic beyond RNA polymerase. *Nat. Rev.*
650 *Microbiol.* **10,** 203–212 (2012).

651 59. Wigneshweraraj, S. *et al.* Modus operandi of the bacterial RNA polymerase containing the
652 sigma54 promoter-specificity factor. *Mol. Microbiol.* **68,** 538–546 (2008).

653 60. Jurado, P., Fernández, L. A. & de Lorenzo, V. Sigma 54 levels and physiological control of
654 the Pseudomonas putida Pu promoter. *J. Bacteriol.* **185,** 3379–3383 (2003).

655 61. Ronneau, S. & Hallez, R. Make and break the alarmone: regulation of (p)ppGpp
656 synthetase/hydrolase enzymes in bacteria. *FEMS Microbiol. Rev.* **43,** 389–400 (2019).

657 62. Bernardo, L. M. D., Johansson, L. U. M., Solera, D., Skärfstad, E. & Shingler, V. The
658 guanosine tetraphosphate (ppGpp) alarmone, DksA and promoter affinity for RNA polymerase in
659 regulation of sigma-dependent transcription. *Mol. Microbiol.* **60,** 749–764 (2006).

660 63. Potvin, E., Sanschagrin, F. & Levesque, R. C. Sigma factors in Pseudomonas aeruginosa.
661 *FEMS Microbiol. Rev.* **32,** 38–55 (2008).

662 64. Shingler, V. Signal sensory systems that impact σ^{54} -dependent transcription. *FEMS Microbiol.*
663 *Rev.* **35,** 425–440 (2011).

664 65. Bellieny-Rabelo, D., Nkomo, N. P., Shyntum, D. Y. & Moleleki, L. N. Horizontally Acquired
665 Quorum-Sensing Regulators Recruited by the PhoP Regulatory Network Expand the Host
666 Adaptation Repertoire in the Phytopathogen Pectobacterium brasiliense. *mSystems* **5,** (2020).

667 66. Linsky, M., Vitkin, Y. & Segal, G. A Novel Legionella Genomic Island Encodes a Copper-
668 Responsive Regulatory System and a Single Icm/Dot Effector Protein Transcriptionally Activated
669 by Copper. *MBio* **11,** (2020).

670 67. Price, M. N., Dehal, P. S. & Arkin, A. P. Horizontal gene transfer and the evolution of
671 transcriptional regulation in Escherichia coli. *Genome Biol.* **9,** R4 (2008).

672 68. Liu, Y. *et al.* Transcriptomic dissection of the horizontally acquired response regulator EsrB
673 reveals its global regulatory roles in the physiological adaptation and activation of T3SS and the
674 cognate effector repertoire in Edwardsiella piscicida during infection toward turbot. *Virulence* **8,**
675 1355–1377 (2017).

676 69. Wu, X. *et al.* Comparative genomics and functional analysis of niche-specific adaptation in
677 Pseudomonas putida. *FEMS Microbiol. Rev.* **35,** 299–323 (2011).

678 70. Treangen, T. J. & Rocha, E. P. C. Horizontal transfer, not duplication, drives the expansion of
679 protein families in prokaryotes. *PLoS Genet.* **7,** e1001284 (2011).

680 71. Smanski, M. J. *et al.* Synthetic biology to access and expand nature's chemical diversity. *Nat.*
681 *Rev. Microbiol.* **14,** 135–149 (2016).

682 72. Way, J. C., Collins, J. J., Keasling, J. D. & Silver, P. A. Integrating biological redesign: where
683 synthetic biology came from and where it needs to go. *Cell* **157,** 151–161 (2014).

684 73. Maervoet, V. E. T. & Briers, Y. Synthetic biology of modular proteins. *Bioengineered* **8,** 196–
685 202 (2017).

686 74. Barajas, J. F., Blake-Hedges, J. M., Bailey, C. B., Curran, S. & Keasling, J. D. Engineered
687 polyketides: Synergy between protein and host level engineering. *Synthetic and Systems*
688 *Biotechnology* **2,** 147–166 (2017).

689 75. Rhodius, V. A. *et al.* Design of orthogonal genetic switches based on a crosstalk map of σs,
690 anti-σs, and promoters. *Mol. Syst. Biol.* **9,** 702 (2013).

691 76. Schmidl, S. R. *et al.* Rewiring bacterial two-component systems by modular DNA-binding
692 domain swapping. *Nat. Chem. Biol.* **15,** 690–698 (2019).

693 77. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search:
694 HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41,** e121 (2013).

695 78. Gumerov, V. M., Ortega, D. R., Adebali, O., Ulrich, L. E. & Zhulin, I. B. MiST 3.0: an updated
696 microbial signal transduction database with an emphasis on chemosensory systems. *Nucleic*
697 *Acids Res.* **48,** D459–D464 (2020).

698 79. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
699 improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).

700 80. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 — approximately maximum-likelihood
701 trees for large alignments. *PLoS ONE* **5,** e9490 (2010).

702 81. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of
703 phylogenomic data. *Mol. Biol. Evol.* **33,** 1635–1638 (2016).

704 82. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases.
705 *Nat. Methods* **6,** 343–345 (2009).

706 83. Hillson, N. J., Rosengarten, R. D. & Keasling, J. D. j5 DNA assembly design automation
707 software. *ACS Synth. Biol.* **1,** 14–21 (2012).

708 84. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein*
709 *Expr. Purif.* **41,** 207–234 (2005).

710 85. Chiniquy, J., Garber, M. E., Mukhopadhyay, A. & Hillson, N. J. Fluorescent amplification for
711 next generation sequencing (FA-NGS) library preparation. *BMC Genomics* **21,** 85 (2020).

712 86. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
713 data. *Bioinformatics* **30,** 2114–2120 (2014).

714 87. Silva, G. G. Z., Cuevas, D. A., Dutilh, B. E. & Edwards, R. A. FOCUS: an alignment-free model
715 to identify organisms in metagenomes using non-negative least squares. *PeerJ* **2,** e425 (2014).

716 88. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient
717 alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).

718 89. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–
719 2079 (2009).

720 90. Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq
721 experiments for DNA-binding proteins. *Nat. Biotechnol.* **26,** 1351–1359 (2008).

722 91. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*
723 **37,** W202-8 (2009).

724 92. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative
725 genomics. *Genome Biol.* **20,** 238 (2019).

726 93. Mironov, A. A., Vinokurova, N. P. & Gelfand, M. S. Software for analysis of bacterial genomes.
727 *Mol Biol (NY)* **34,** 222–231 (2000).

728 94. Thompson, M. G. *et al.* Massively Parallel Fitness Profiling Reveals Multiple Novel Enzymes
729 in Pseudomonas putida Lysine Metabolism. *MBio* **10,** (2019).

730 95. Rand, J. M. *et al.* A metabolic pathway for catabolizing levulinic acid in bacteria. *Nat. Microbiol.*
731 **2,** 1624–1634 (2017).

732 96. Incha, M. R. *et al.* Leveraging host metabolism for bisdemethoxycurcumin production in
733 Pseudomonas putida. *Metab. Eng. Commun.* **10,** e00119 (2020).

734 97. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with
735 high throughput capability. *PLoS ONE* **3,** e3647 (2008).

736 98. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression
737 and hybridization array data repository. *Nucleic Acids Res.* **30,** 207–210 (2002).

738

739

740