Experimental evolution-driven identification of Arabidopsis rhizosphere competence genes in *Pseudomonas protegens*

Erqin Li[1,#], Henan Jiang[1], Corné M.J. Pieterse[1], Alexandre Jousset[2], Peter A.H.M. Bakker[1], Ronnie de Jonge[1]

[1]Plant-Microbe Interactions, Department of Biology, Science4Life, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands

[2]Ecology and Biodiversity, Department of Biology, Science4Life, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands.

[†]Corresponding author: Ronnie de Jonge; r.dejonge@uu.nl

#Current address: Freie Universität Berlin, Institut für Biologie, Dahlem Center of Plant Sciences, Plant Ecology, Altensteinstr. 6, D-14195, Berlin, Germany, erqinli22@gmail.com

Abstract

Beneficial plant root-associated microorganisms carry out a range of functions that are essential for plant performance. Establishment of a bacterium on plant roots, however, requires overcoming several challenges, including the ability to outcompete neighboring microorganisms and suppression of plant immunity. Forward and reverse genetics approaches have led to the identification of diverse mechanisms that are used by beneficial microorganisms to overcome these challenges such as the production of iron-chelating compounds, biofilm formation, or downregulation of plant immunity. However, how such mechanisms have developed from an evolutionary perspective is much less understood. In an attempt to study bacterial adaptation in the rhizosphere, we employed an experimental evolution approach to track the physiological and genetic dynamics of root-dwelling *Pseudomonas protegens* CHA0 in the *Arabidopsis thaliana* rhizosphere under axenic conditions. This simplified binary one plant, and one bacterium system allows for the amplification of key adaptive mechanisms for bacterial rhizosphere colonization. We found that mutations in global regulators, as well as in genes for siderophore production, cell surface decoration, attachment, and motility accumulated in parallel in our evolutionary experiment, underlining several different strategies of bacterial adaptation to the rhizosphere. In total we identified 35 mutations, including single-nucleotide polymorphisms, smaller indels and larger deletions, distributed over 28 genes in total. Altogether these results underscore the strength of experimental evolution to identify key genes and pathways for bacterial rhizosphere colonization, as well as highlighting a methodology for the development of elite beneficial microorganisms with enhanced root-colonizing capacities that can support sustainable agriculture in the future.

35 ## Introduction

36 Plants are associated with complex microbial communities assembled into a functional microbiome that
37 safeguards optimal plant performance under harsh environmental conditions (1). The rhizosphere
38 microbiome is a particularly interesting hotspot of plant-microbe interactions. Plants deposit up to 44%
39 of their photosynthetically fixed carbon into the rhizosphere, fueling a specific microbial community (2).
40 The microbial species pool in the bulk soil is the source from which the root microbiome is recruited, and
41 plant genotype, immune responses, and environmental factors are postulated to affect this process (3–
42 6). The establishment of beneficial microbial associations requires a high degree of coordination of both
43 plant and microbial responses by means of a continuous molecular dialogue (7, 8). Plant-associated
44 microorganisms can improve plant yield by protecting the plant from abiotic stresses (9), improving plant
45 nutrition and growth (10–12), antagonizing soil-borne pathogens (13), or stimulating plant immunity (14).
46 To exert their beneficial effects on plant performance, bacteria must colonize the roots efficiently and
47 establish significant populations. For example, population densities above 10^5 cells per gram of root are
48 required for efficient suppression of soil-borne plant pathogens by *Pseudomonas* spp. (15, 16). Therefore,
49 bacterial adaptation to the plant root environment may be essential for the successful implementation of
50 microbiome services in agriculture in order to support plant health.

51 Among all root-dwelling organisms, fluorescent *Pseudomonas* spp. are well characterized in terms of the
52 traits required for their growth in the rhizosphere and for the establishment of beneficial plant-microbe
53 interactions (17). To study bacterial traits involved in efficient root colonization, mutants defective in
54 specific traits suspected to be involved in colonization, are compared to the ability of the parental strain
55 to colonize roots. Such studies have highlighted a range of bacterial traits involved in efficient root
56 colonization, including flagella (18), surface lipopolysaccharides (19), and amino acid synthesis (20). Using
57 random mutagenesis and by determining the fitness of each mutant in competition with its parental strain
58 in the rhizosphere, many other traits important for rhizosphere competence in *Pseudomonas* were
59 discovered (17). Recently, random mutagenesis in *Pseudomonas capeferrum* WCS358 led to the
60 identification of two genes important for gluconic acid biosynthesis which in turn is essential for the
61 suppression of local, flagellin-induced root immune responses (21, 22). Such suppression was shown to
62 be important for rhizosphere competence as gluconic acid-deficient mutants maintain reduced
63 rhizospheric populations (22). In another recent study, genome-wide saturation mutagenesis of
64 *Pseudomonas simiae* WCS417r revealed that 2% of the protein-coding genes are important for successful
65 root colonization (23). Mutations that negatively affect rhizosphere competence, and mutations that

66 confer a competitive advantage were identified in this study. The observation that mutations can lead to
67 increased root colonization (23), suggests that there is room for improvement of bacterial fitness in the
68 rhizosphere.

69 In the present study we used an experimental evolution approach (24) to study how *Pseudomonas*
70 *protegens* CHA0 (CHA0) evolves during repeated colonization of the rhizosphere of the model plant
71 *Arabidopsis thaliana* (Arabidopsis). The model biological control agent CHA0 displays broad-spectrum
72 antagonistic activity against several plant pathogenic fungi and bacteria (11), and its complete genome
73 sequence is available (25). We performed highly controlled experimental evolution in a gnotobiotic and
74 carbon-free sand system in which bacteria depend solely on the plant for supply of carbon. Following
75 inoculation and establishment on the roots, bacterial populations were transferred to new plants, and
76 this cycle was repeated eight times. We hypothesized that the repeated colonization and establishment
77 of the bacterial population on the plant root would create an environment in which selection pressure
78 drives the accumulation of better colonizers. *In vitro* characterization of individual bacterial colonies from
79 these populations combined with sequencing analysis led to the identification of several evolutionary
80 trajectories involving 35 distinct mutations that impact social traits representing inter-population
81 communication and cooperation, carbon source utilization, motility, or biocontrol activity. By combining
82 experimental evolution with whole genome re-sequencing, we created a powerful screen for the
83 identification of adaptive mutations with positive effects on rhizosphere colonization.

84

85

86 ## Results

87 ### Mutational events in independent evolutionary lines

88 We previously studied five experimental evolutionary populations, referred to as lines, of CHA0 evolving
89 in the rhizosphere of Arabidopsis in a gnotobiotic system. Independent populations were introduced on
90 the roots and after four weeks of plant growth the populations were transferred to new plants. This cycle
91 of transferring was repeated eight times (26). To study adaptation at the individual level, here, we isolated
92 in total 90 bacterial colonies at the end of plant growth cycles 2, 4 and 6. These colonies, and six colonies
93 from the ancestral population that was initially introduced on the roots, were re-sequenced to an average
94 depth of 25-fold coverage (minimum 10, maximum 70) and used for the identification of single nucleotide
95 polymorphisms (SNPs), as well as small and large insertions or deletions (INDELs). The six ancestral

96 colonies were all identical, indicating that there was no genetic variation in the starting population and all
97 observed mutations are therefore *de novo* mutations. In total, one or more mutations were detected in
98 64 evolved colonies, which collectively represent 71% of the 90 characterized bacterial colonies (Table
99 S1). We identified 5 synonymous substitutions, 20 nonsynonymous substitutions, and 4 deletions ranging
100 in length from 1 base pair (bp) to about 400 bp, distributed over 22 genes, and 6 additional mutational
101 events located in intergenic regions (Table 1). Mutations located in intergenic regions possibly affect
102 transcription of nearby genes via affecting their regulatory proteins binding sites and subsequent changes
103 in their promoter activity (27, 28). Several mutations were found to be clustered in select genes and/or
104 regions on the CHA0 genome, e.g. those targeting the response regulator *gacA* (PFLCHA0_RS17965;
105 *NC_021237:4,039,113-4,039,754*) or the *OBC3* gene cluster involved in lipopolysaccharide (LPS)
106 biosynthesis (*NC_021237:2,173,707-2,196,028*)(29), but the majority of the mutations are spread across
107 the 6.1 Mbp CHA0 genome (Figure 1). Functional characterization of the mutated genes by analysis of
108 their Clusters of Orthologous Groups (COGs) annotation revealed that the majority of these genes are
109 involved in transcription (COG term 'K'), signal transduction mechanisms (COG term 'T'), amino acid
110 transport and metabolism (COG term 'E') and cell wall/membrane/envelope biogenesis (COG term 'M')
111 (Figure 1, Table 1).

112 **Table 1.** Mutations that occurred in 28 genes during Arabidopsis rhizosphere adaptation divided over five replicate CHA0 populations (lines)

Locus tag	Gene name	Description	COG[1]	Line	Total number of alleles	Number of mutations in protein-coding genes			Number of intergenic mutations
						Non-synonymous	Synonymous	Deletions	
PFLCHA0_RS02080	*hutI*	Imidazolonepropionase	F	4	1	0	1	0	0
PFLCHA0_RS03400	*accC*	Biotin carboxylase, acetyl-CoA carboxylase subunit	I	2	1	1	0	0	0
PFLCHA0_RS05510	*nudL*	Coenzyme A pyrophosphatase/nudix hydrolase NudL	L	4	1	0	1	0	0
PFLCHA0_RS06125	*rpoS*	RNA polymerase sigma factor RpoS	K	5	1	1	0	0	0
PFLCHA0_RS08340	*fleQ*	Sigma-54-dependent Fis family transcriptional regulator	T	5	1	1	0	0	0
PFLCHA0_RS08490	*flhA*	Flagellar biosynthesis protein FlhA	N	4	1	0	0	1	0
PFLCHA0_RS09880	*RS09880*	Glycosyltransferase (GT)	M	5	1	1	0	0	0
PFLCHA0_RS09890	*oafA*	O-antigen acetylase	I	1, 3	2	1	0	1	0
PFLCHA0_RS09920	*galE*	UDP-glucose 4-epimerase	M	2	1	1	0	0	0
PFLCHA0_RS11785	*RS11785*	LysR family transcriptional regulator	K	4	1	1	0	0	0
PFLCHA0_RS11820	*RS11820*	PaaI family thioesterase	Q	4	1	0	1	0	0
PFLCHA0_RS12070	*RS12070*	2OG-Fe(II) oxygenase superfamily	S	5	1	0	1	0	0
PFLCHA0_RS13000	*yvaQ2*	Methyl-accepting chemotaxis protein	NT	2	1	0	0	0	1
PFLCHA0_RS14960	*tetR*	TetR/AcrR family transcriptional regulator	K	5	1	1	0	0	0
PFLCHA0_RS17350	*RS17350*	Methyltransferase domain-containing protein	H	1	1	0	0	1	0
PFLCHA0_RS17965	*gacA*	UvrY/SirA/GacA family response regulator transcription factor	K	1, 2, 4	6	5	0	0	1
PFLCHA0_RS18525	*RS18525*	ABC transporter substrate-binding protein	E	5	1	1	0	0	0
PFLCHA0_RS21265	*pvdS*	RNA polymerase factor sigma-70	K	1	1	0	0	0	1
PFLCHA0_RS21275	*RS21275*	Transporter substrate-binding domain-containing protein	ET	1	1	1	0	0	0
PFLCHA0_RS21855	*wbpM*	Polysaccharide biosynthesis protein/NDP-sugar epimerase	GM	1	1	1	0	0	0
PFLCHA0_RS22600	*gacS*	Hybrid sensor histidine kinase/response regulator	T	3	1	1	0	0	0
PFLCHA0_RS22950	*argT5*	ABC transporter substrate-binding protein/Lysine-arginine-ornithine-binding periplasmic protein ArgT	ET	4	1	0	0	1	0
PFLCHA0_RS25175	*mraZ*	Division/cell wall cluster transcriptional repressor MraZ	K	2	1	0	0	0	1
PFLCHA0_RS26215	*osmY*	Osmotically-inducible protein OsmY/BON domain-containing protein	S	2	1	0	0	0	1
PFLCHA0_RS27515	*rpsH*	30S ribosomal protein S8	J	4	1	0	0	0	1
PFLCHA0_RS30075	*sadB*	Surface attachment defective (SadB) ortholog/HDOD domain-containing protein	T	1	2	2	0	0	0
PFLCHA0_RS30120	*tssM*	Type VI secretion system membrane subunit TssM	S	2	1	0	1	0	0
PFLCHA0_RS31060	*nlpD*	Lipoprotein nlpD/lppB/LysM domain-containing protein	D	3	1	1	0	0	0
Total	28				35	20	5	4	6

113 [1]Clusters of Orthologous Groups (COGs); https://www.ncbi.nlm.nih.gov/COG/(30)

114 **Identification of potential root colonization genes**

115 Bacterial genes that are involved in colonization of plant roots can be revealed by identifying beneficial
116 mutations that evolve during adaption of bacteria to the rhizosphere environment and have positive
117 effects on root colonization. Over time, such mutants will outcompete the ancestral strain and become
118 dominant in the bacterial population. The observation that only a limited number of mutations
119 accumulated relative to the total number of genes within the genome across the entire experiment,
120 makes it highly unlikely that the same gene would acquire several changes by chance in independent
121 evolutionary lines. Nevertheless, we observe recurrent mutations in several of the same genes and/or
122 pathways (Table 1), which is a strong indication for adaptive evolution. Genes or pathways that acquired
123 mutations in multiple independent CHA0 populations included *gacA* and *gacS*, the *OBC3* gene cluster
124 (*oafA*; *galE*; *PFLCHA0_RS09880* (29)), and a putative pyoverdine siderophore biosynthesis cluster (*pvdS*;
125 *PFLCHA0_RS21275*). Other genes were targeted more than once in the same population, such as *sadB*.
126 Because these genes were repeatedly identified in the CHA0 evolution experiment, they can be assumed
127 to contribute significantly to bacterial fitness in the rhizosphere (Table 1, Figure 2). Moreover, it suggests
128 that the independent evolutionary lines converge on similar evolutionary trajectories involving
129 overlapping biological processes and molecular mechanisms.

130 To illustrate the diverse evolutionary trajectories, we constructed phylogenetic trees of each replicate
131 population (line) using all mutations that were detected (Figure 3). In each line, both the number and
132 depth of branches are very different (Figure 3, Table 2). In line 1 and 3, the populations were swept early
133 by *oafA* mutants and later by specific *gac* mutants, while in line 2 and 4, co-existence of multiple
134 genotypes characterized by *gac* mutations was observed. The evolved isolates contain one to four
135 mutations in general, and in each line, 3 (line 3) to 9 (line 2 and 4) mutations were identified in total. All
136 mutations were unique for each line, as is expected for independently evolving populations.

137 The frequency of each mutation (Table 2) was determined from the six bacterial colonies that were
138 isolated and sequenced from each evolutionary line at cycles 2, 4 and 6. To further investigate and verify
139 the changes in mutation frequency in the entire CHA0 population at the end of a given cycle, we set out
140 to determine the frequency of three of the mutations, i.e. $gacA^{D49Y}$, $gacA^{D54Y}$, and $gacS^{G27D}$, with increased
141 accuracy and resolution during long-term rhizosphere adaptation. To this end, we developed a PCR-based
142 high-resolution melting (HRM) analysis using mutation-specific HRM probes (Table S2), that allow for the
143 accurate quantification of mutant frequency across a wide range for all three mutations ($p<0.001$; Figure
144 S1). Using this method, we found that the population-level frequency of these three mutations in the

145 three respective evolutionary lines at the end of each of the eight rhizosphere adaptation cycles
146 corresponded well with the frequency previously obtained from the six cultured and sequenced isolates
147 (Table 2, Figure S2). These findings indicate that the culture-based quantification of mutant frequency
148 provides a realistic measure for population-level frequency. Moreover, it can be seen that whereas the
149 *gacS*^{G27D} mutation reaches fixation after eight cycles of rhizosphere adaptation, the *gacA*^{D49Y} and *gacA*^{D54Y}
150 mutants stabilize at around 50% and 25%, respectively, indicative of frequency-dependent (FD) selection
151 (Figure S2). FD selection describes the phenomenon that the fitness of a particular genotype or phenotype
152 is dependent on its frequency. Such context-dependency has been frequently linked to cheating
153 behaviour in which microbial cells that lack or limit the production of certain costly compounds benefit
154 from other cells that do produce these compounds. When a minimal amount of such compound increases
155 genotype or phenotype fitness, FD selection can occur resulting at the stabilization of the mutation
156 frequency.

157

158 **Table 2.** Nomenclature for CHA0 mutations and variants and their frequency/cycle in each respective population.

Line	Locus tag	Gene name	DNA sequence change[1]	Amino acid change (three-letter code)[2]	Standard nomenclature (used in text)	Frequency in the population (x/6)			
						Cycle 2	Cycle 4	Cycle 6	RQ-HRM[3]
1	PFLCHA0_RS09890	*oafA*	c.1005C>A	p.Tyr335X	Y335X	4/6	6/6	6/6	
1	PFLCHA0_RS17350	*RS17350*	c.230delC	p.Ala77Ala.fsX14	A77A.fsX14	2/6	6/6	6/6	
1	PFLCHA0_RS17965	*gacA*	c.145G>T	p.Asp49Tyr	D49Y	-	2/6	4/6	x
1	PFLCHA0_RS21855	*wbpM*	c.235G>C	p.Gly79Arg	G79R	-	1/6	-	
1	PFLCHA0_RS21275	*RS21275*	c.233G>T	p.Arg78Leu	R78L	-	1/6	-	
1	PFLCHA0_RS21265	*pvdS*	c.-33T>C	NA	-33T>C	-	-	1/6	
1	PFLCHA0_RS30075	*sadB*	c.548G>C	p.Arg183Pro	R183P	-	-	1/6	
1			c.773T>A	p.Leu258Gln	L258Q	-	-	2/6	
2	PFLCHA0_RS17965	*gacA*	c.548A>C	p.Tyr183Ser	Y183S	2/6	1/6	-	
2	PFLCHA0_RS09920	*galE*	c.94G>A	p.Val32Met	V32M	2/6	5/6	6/6	
2	PFLCHA0_RS30120	*tssM*	c.1336C>T	p.Leu446Leu	1336C>T	1/6	-	-	
2	PFLCHA0_RS03400	*accC*	c.1237G>A	p.Glu413Lys	E413K	-	4/6	6/6	
2	PFLCHA0_RS13000	*yvaQ2*	c.-9G>T	NA	-9G>T	-	1/6	-	
2	PFLCHA0_RS17965	*gacA*	c.160G>T	p.As54Tyr	D54Y	-	-	2/6	x
2	PFLCHA0_RS17965	*gacA*	c.289G>A	p.Gly97Ser	G97S	-	-	3/6	
2	PFLCHA0_RS25175	*mraZ*	c.-211A>G	NA	-211A>G	-	-	2/6	
2	PFLCHA0_RS26215	*osmY*	c.-104G>A	NA	-104G>A	-	-	1/6	
3	PFLCHA0_RS09890	*oafA*	c.1013delA	p.Lys338Ser.fsX18	K338S.fsX18	1/6	5/6	6/6	
3	PFLCHA0_RS22600	*gacS*	c.80G>A	p.Gly27Asp	G27D	-	1/6	4/6	x
3	PFLCHA0_RS31060	*nlpD*	c.590A>C	p.Gln197Pro	Q197P	-	1/6	-	
4	PFLCHA0_RS22950	*argT5*	c.120-533del	p.Lys40-Leu185del.fsX6	Δ*argT5*	2/6	-	-	
4	PFLCHA0_RS02080	*hutI*	c.786C>T	p.Phe262Phe	786C>T	1/6	-	-	
4	PFLCHA0_RS17965	*gacA*	c.-40T>A	NA	-40T>A	-	2/6	1/4	
4	PFLCHA0_RS17965	*gacA*	c.112G>T	p.E38X	E38X	-	2/6	3/4	
4	PFLCHA0_RS05510	*nudL*	c.288C>T	p.Ser96Ser	288C>T	-	1/6	-	
4	PFLCHA0_RS11820	*RS11820*	c.33C>T	p.Ala11Ala	33C>T	-	1/6	-	
4	PFLCHA0_RS08490	*flhA*	c.1176delT	p.His393Gln.fsX15	H393Q.fsX15	-	-	3/4	
4	PFLCHA0_RS11785	*RS11785*	c.766A>T	p.Ser256Cys	S256C	-	-	3/4	
4	PFLCHA0_RS27515	*rpsH*	c.-90C>G	NA	-90C>G	-	-	1/4	
5	PFLCHA0_RS18525	*RS18525*	c.254delT	p.Val85Gly.fsX40	V85G.fsX40	1/6	-	-	
5	PFLCHA0_RS06125	*rpoS*	c.193C>T	p.Gln65X	Q65X	-	1/6	4/6	
5	PFLCHA0_RS08340	*fleQ*	c.959G>A	p.Arg320Gln	R320Q	-	-	1/6	
5	PFLCHA0_RS14960	*tetR*	c.381C>G	p.Tyr127X	Y127X	-	-	2/6	
5	PFLCHA0_RS09880	*RS09880*	c.246C>A	p.Tyr82X	Y82X	-	-	1/6	
5	PFLCHA0_RS12070	*RS12070*	c.1389C>G	p.Pro463Pro	1389C>G	-	-	1/6	

159 [1]DNA sequence change positions are relative to the cDNA (c); del represents a deletion

160 [2]X represents a stop codon (at its relative position in case of a shifted frame); fs represents a frame shift; del represents a deletion

161 [3]Mutations that are selected for the analysis of their frequency trajectories using RQ-HMR analysis

162 **Early adaptations are driven by cell surface-related genes**

163 In general, mutations that are fixed early on in the rhizosphere adaptation process tend to have a high
164 selective advantage (31). Disruptive mutations in *oafA*, resulting in premature stops halfway the coding
165 region, were detected as the first acquired mutations in two independent evolutionary lines (lines 1 and
166 3) and appear to have swept the population in later generations (Table 2). This gene, which is part of the
167 previously described *OBC3* gene cluster (29), encodes an O-acetyltransferase, which is postulated to
168 acetylate the O-antigen component of the outer membrane LPS (32). Another *OBC3* mutation that
169 accumulated early on in the rhizosphere adaptation process was $galE^{c.94G>A}$ leading to an amino acid
170 substitution (V32M) in galE which swept through the population in evolutionary line 2, reaching fixation
171 in cycle 6 (Table 2). *GalE* encodes an UDP-glucose 4-epimerase which is involved in O-antigen and LPS core
172 biosynthesis (33–35). One colony with a mutation in a third *OBC3* cluster gene, *RS09880*, encoding a
173 putative glycosyl transferase was found in evolutionary line 5. Thus, in four out of the five evolutionary
174 lines mutations that likely affect bacterial LPS structure appeared during rhizosphere adaptation, and in
175 three evolutionary lines such mutations even became dominant in the population. These results strongly
176 suggest that modifying bacterial cell surface structure is an important bacterial strategy in early
177 adaptation to the rhizosphere.

178

179 **Adaption driven by global regulators**

180 In the present study, six mutations were detected in *gacA* in three of the five evolutionary lines,
181 representing approximately 20% of all missense mutations. Notably, in evolutionary lines 2 and 4, multiple
182 *gacA* alleles accumulated, some of which were detected in early plant adaptation cycles (Table 2).
183 Additionally, a *gacS* mutation accumulated in line 3. These two genes encode the main constituents of the
184 conserved GacA/GacS two-component regulator system, i.e. the hybrid sensor histidine kinase GacS and
185 the cognate response regulator GacA (Figure 2). In gram-negative bacteria, activation of GacS results in
186 cross-phosphorylation of GacA via phosphotransfer which in turn leads to activation of the expression of
187 the small RNA genes *rsmY* and *rsmZ* via its helix-turn-helix (HTH) domain-binding domain (36). In CHA0,
188 this regulatory pathway is known to control quorum sensing as well as secondary metabolism and stress
189 resistance (37–40). In the closely related strain *P. protegens* Pf-5 this pathway was shown to have a big
190 impact on bacterial genome-wide gene expression, affecting the expression of more than 10% of the
191 annotated genes (41). Similarly, *gacA* mutants that arose on the edge of swarming Pf-5 colonies showed
192 dramatically altered genome-wide gene expression patterns (42).

193 Including *gacS* and *gacA*, about half of the mutated genes in this study are global regulators or sigma
194 factors (Table 1, Figure 1). This high frequency suggests that global regulator-controlled networks are
195 evolvable and play a major role in rapid bacterial adaptation. Pleiotropic adaptive mutations in global
196 regulator genes have been shown to be important for bacterial adaption both in laboratory (43, 44), and
197 in natural settings (45, 46). Remodeling and continuous optimization of existing regulatory networks by
198 single mutations is an important strategy for bacterial adaptation to the host (47).

199

200 **Bacterial motility**

201 Bacterial motility is an important trait for rhizosphere competence, mediating colonization of distal parts
202 of the root system (48), and both LPS and the GacS/GacA two-component regulator system affect this
203 trait. The O-antigenic side chain of the LPS was reported to contribute to swimming and swarming motility
204 in the plant pathogenic bacterium *Erwinia amylovora* (49, 50). The GacS/GacA two-component regulator
205 system controls bacterial motility, for example by affecting transcription of genes related to flagella and
206 biosurfactant biosynthesis (42, 48, 51). We also identified several other mutations across the various
207 evolutionary lines that can be linked to bacterial motility. For instance, a disruptive mutation in the
208 flagellar biosynthesis protein-coding gene *FlhA* (H393Q.fsX15) that is involved in the biogenesis of the
209 bacterial flagellum (Table 1) appeared in evolutionary line 4 reaching up to a predicted frequency of 75%
210 (Table 2). In *P. fluorescens* Pf-5, and in *P. aeruginosa* PA01 FlhA is reported to be essential for swimming
211 motility (41).

212 Furthermore, we identified one amino-acid substitution, R320Q, in FleQ and two in SadB, R183P and
213 L258Q respectively, that based on sequence similarity to well-studied, homologous proteins in other
214 bacteria are likely to affect motility as well. FleQ is a σ^{54}-dependent Fis family transcriptional regulator
215 which regulates flagellar motility, biofilm formation as well as Pel exopolysaccharide production in
216 response to cellular c-di-GMP in *P. aeruginosa* (52). *P. protegens* FleQ shares 84% sequence identity and
217 98% sequence coverage with *P. aeruginosa* FleQ and similarly contains an N-terminal flagellar regulatory
218 domain FleQ (PF06490) also known as the REC receiver domain, a central AAA+/ATPase σ^{54}-interaction
219 domain (PF00158), and a C-terminal Fis-type HTH DNA-binding domain (PF02954). The R320Q substitution
220 we identified here is found in the AAA+/ATPase σ^{54}-interacting domain in between the arginine (R) finger
221 (300-303) and a c-di-GMP-binding motif, ExxxR (330-334) (52). The conserved arginine residues in FleQ,
222 including the here mutated R320, are thought to be important for protein oligomerization, and
223 substitution of any of them abolishes ATPase activity in *Vibrio cholerae* EpsE completely (53). SadB, finally,

224 encodes a HDOD domain-containing protein (Table 1) which shares 77% sequence identity and 99%
225 sequence coverage with *P. aeruginosa* SadB (PA5346). In *P. aeruginosa*, SadB stimulates Pel
226 exopolysaccharide production and the chemotaxis-like cluster CheIV, which in turn affect flagellar motility
227 as well as biofilm formation (54). In *Pseudomonas fluorescens* F113, SadB, together with FleQ control
228 flagellar motility, dependent and independent of the GacS/GacA two-component regulator system (55,
229 56).

230 Parallelism of targeted mutations on this functional motility pathway impelled us to assess bacterial
231 motility and track its dynamics across multiple evolutionary lines. Two independent evolutionary
232 trajectories to genotypes *galE*^{V32M} · *accC*^{E413K} · *gacA*^{D54Y} and *oafA*$^{K338S.fsX18}$ · *gacS*^{G27D} from evolutionary lines
233 2 and 3, respectively, were studied for changes in their swimming and swarming behavior. Both genotypes
234 accumulated mutations in *OBC3* cluster genes (29) linked to LPS biogenesis (*galE* or *oafA*) and the
235 GacS/GacA two-component regulator system (*gacA* or *gacS*). Interestingly, *galE*^{V32M} and *oafA*$^{K338S.fsX18}$
236 mutants displayed opposing phenotypes when their swimming and swarming behavior was examined.
237 Whereas the *galE*^{V32M} mutant produces significantly larger swimming and swarming haloes as compared
238 to the wild type, the *oafA*$^{K338S.fsX18}$ mutant produces significantly smaller swimming haloes and is even
239 completely defective in swarming (Figure 4). The *gacA* and *gacS* mutations, however, consistently reduced
240 swimming and swarming behavior.

241

242 **Dynamics of global phenotypic change**

243 Since natural selection eventually operates at the phenotypic level, revealing bacterial global phenotypic
244 evolutionary dynamics can help us to identify traits that are under selection. Moreover, beneficial genetic
245 mutations can be predicted if they are linked to well-known root colonization traits. A broad range of 30
246 bacterial traits including different aspects of bacterial life-history traits, were measured for the sequenced
247 isolates, allowing genotype-phenotype association analysis.

248 As shown in Figure 5a, the 30 bacterial traits separated into four clusters that share a similar pattern
249 across the different mutant genotypes and this clustering is supported by model-based clustering analysis
250 (Figure S3). Growth of the bacteria in King's B (KB) medium was positively correlated with siderophore
251 and exoprotease production, tryptophan side oxidase activity, and growth inhibition of the bacterial plant
252 pathogen *Ralstonia solanacearum*. Thus, this cluster, designated cluster 1, contains traits associated with
253 bacterial social behavior, related to within population communication and cooperation, such as the

254 production of public goods. In a principle component analysis (Figure S4a), the first principle component
255 (PC1) is strongly correlated with all five traits, with a total explanation of 64.5% for all variables. Cluster 2
256 contains traits linked to carbon source utilization. For this cluster PC1 is strongly correlated with all carbon
257 source usage-related traits, with a total explanation of 83.9% for all variables (Figure S4b). A third cluster
258 was observed for the bacterial ability to form a biofilm, to produce indole-3-acetic acid (IAA), and to inhibit
259 the growth of two fungal plant pathogens, with a total explanation of 82.4% by PC1 (Figure S4c). Finally,
260 cluster 4 contains all seven traits that are related to bacterial resistance to biotic and abiotic stresses. The
261 first principle component (PC1) is strongly correlated with all seven traits, with a total explanation of
262 53.9% for all variables (Figure S4d). For these four clusters, the PC1 (or -PC1) value is used as a proxy to
263 present the general performance of all the traits that clustered together.

264 As is shown in Figure 5b, all five evolutionary lines showed a parallel trend of accumulative decline of
265 social traits. Mutations in *gacA, gacS,* and *rpoS* resulted in significantly decreased bacterial social traits,
266 and this was also observed for the double mutant *RS11785*^{S256C} · *flhA*$^{H393Q.fsX15}$ relative to the background
267 *gacA*^{E38X} mutation. *RS11785*, like *gacA*, encodes a LysR-type transcriptional regulator. In addition to these
268 global regulators, earlier mutations in *oafA*, *galE* and *RS17350*$^{A77A.fsX14}$ encoding a putative
269 methyltransferase, also resulted in a significant but relatively small decrease in social traits. This parallel
270 decline of the bacterial social traits index suggests a negative selection of bacterial social behavior,
271 especially the production of costly public goods such as siderophores and exoproteases.

272 All traits related to the utilization of 14 different carbon sources, that were selected based on their
273 reported presence in root exudates of Arabidopsis (57) grouped in one cluster (cluster 2; Figure 5a),
274 suggesting that increase or decrease of the ability to utilize carbon sources is co-regulated. Different
275 mutations in *gacS* and *gacA* resulted in contrasting bacterial carbon source utilization. A mutation in the
276 first N-terminal transmembrane domain of GacS (G27D) (Figure 2) resulted in significant enhancement of
277 carbon source utilization (Figure 5a, c). A similar trend was observed for mutant *gacA*^{G97S}. In contrast, the
278 majority of evolved genotypes, including most mutations in *gacA* and *rpoS*, showed a reduced ability to
279 utilize carbon sources (Figure 5a, c).

280 In contrast to the previous two clusters, traits in clusters 3 (Figure 5d) and 4 (Figure 5e) were more stable,
281 in that most genotypes behave like the ancestor. Only two *gacA* mutants, *gacA*^{E38X} and *gacA*^{Y183S}, were
282 associated with a significant decline of bacterial traits associated with biocontrol, such as antifungal
283 activity and biofilm formation, while the *rpoS*^{Q65X} mutation dramatically increased these traits. The double
284 mutation *RS11785*^{S256C} · *flhA*$^{H393Q.fsX15}$, in the background of the disruptive *gacA*^{E38X} mutation (Figure 5e),

285 was the only mutational event that led to a significant increase of general resistance to various
286 environmental stresses. *FlhA*, as previously mentioned encodes the flagellar biosynthesis protein FlhA and
287 is linked to bacterial motility. The LysR-type transcriptional regulator RS11785 in addition to its N-terminal
288 HTH DNA-binding domain (PF00126) contains a predicted C-terminal substrate binding domain (PF03466),
289 which is linked to aromatic compound degradation and resembles that of the type 2 periplasmic binding
290 proteins or PBP2s. PBP2s are responsible for binding and uptake of several substrates including
291 polysaccharides and several amino acids. We used HHpred (58) for protein remote homology detection
292 and three-dimensional structure analysis to assess the possible consequence of the S256C amino acid
293 substitution. This analysis identified the remote homolog OxyR from *Escherichia coli* as the best template
294 for modeling (E-value: 2.9e-30, Score: 166.45). Intriguingly, it was previously shown that the activity of
295 OxyR, which represents the master peroxide sensor in gram-negative bacteria, is regulated by
296 intramolecular disulfide-bond formation in the regulatory PBP2-like domain. Notably, RS11785 is unlikely
297 to represent the master peroxide sensor in *P. protegens* CHA0, because another CHA0 protein encoded
298 by *PFLCHA0_RS030065* is much more similar (89% percent pairwise sequence identity versus 29%).
299 Nevertheless, it is tempting to speculate that the here observed S256C substitution might influence
300 regulatory activity by altered intramolecular disulfide-bond formation. Lastly, RS11785 bears resemblance
301 (37% sequence identity, 93% sequence coverage) to the LysR-type PBP2-domain containing regulator *alsR*
302 from *Bacillus amyloliquefaciens*. *AlsR* regulates the activity of the acetoin operon (*alsSD*) in response to
303 several signals such as glucose and acetate. Acetoin (3-hydroxy-2-butanone, a precursor for 2,3-
304 butanediol) in turn is known as a potent elicitor of induced systemic resistance (59), and it is linked to
305 general plant growth promotion in *Bacilli*. Given these observations from literature, but in the absence of
306 singular mutations in *flhA* and *RS11785*, the molecular mechanism underlying the observed enhanced
307 resistance in this double mutant remains to be clarified but we hypothesize that altered regulatory activity
308 by RS11785 is a likely driver of the increased stress resistance observed in this mutant.

309 ## Discussion

310 The rhizosphere is a nutrient-rich environment for root-associated bacteria. However, to access the
311 available nutrients bacteria must overcome various challenges, including the plant immune system,
312 presence of competing and/or predatory microorganisms, and abiotic stresses. In the gnotobiotic binary
313 system used in this study we tracked changes in *Pseudomonas protegens* CHA0 in the rhizosphere of
314 Arabidopsis under reproducible and controlled conditions without interference of complex interactions
315 with other microbes. Mutations affecting global regulators, bacterial cell surface structure, and motility
316 accumulated in parallel in our evolutionary experiment, revealing at least three important strategies of
317 bacterial adaptation to the rhizosphere.

318

319 ### Global regulators and rhizosphere adaptation

320 The GacS/GacA two-component regulator system controls the production of antimicrobial secondary
321 metabolites, exoenzymes, and siderophores, but also biofilm formation, stress responses, motility, and
322 quorum sensing (13, 38, 41, 60, 61). In the present study, mutations in the GacS/GacA two-component
323 regulator system caused dramatic changes in several bacterial phenotypic traits, including motility, carbon
324 source utilization, and social traits (Figures 4 and 5). The observed loss of motility does not logically fit
325 with adaptation to the rhizosphere as motility has been reported to be an important root colonization
326 trait in Pseudomonas (17). In line with this, the *gac* mutations that accumulated in *Pseudomonas*
327 *fluorescens* F113 colonizing the rhizosphere of alfalfa showed increased motility and resulted in an
328 enhanced competitive colonization ability (48, 62). Our results fit observations by Song et al (2016)(42) in
329 the closely related *P. protegens* strain Pf-5. During swarming motility, *gac* mutants emerged that lack
330 production of the surfactant orfamide. These mutants cannot swarm, but they co-swarm with orfamide
331 producing cells (42). Thus, the lack of motility in CHA0 may also be the result of loss of surfactant
332 production and leads to a selective advantage in the presence of surfactant producers. Several other
333 mutations that are identified in diverse regulatory, signaling and executionary genes, *flhA*, *fleQ* and *sadB*
334 can be linked to bacterial motility. Although no data is currently available for the *flhA* and *sadB* mutants,
335 preliminary analysis of swarming and swimming capacity *in vitro* shows that the *fleQ* mutant has also lost
336 its ability to move. Possibly, in our experimental system motility is less important than the ability to adhere
337 to the root surface and the formation of a biofilm that is frequently described alongside motility as two
338 sides of the same coin. In many bacteria, these processes in fact appear inversely correlated and

339 controlled by the intracellular secondary messenger cyclic dimeric-GMP (c-di-GMP) (63–65). Future
340 experiments should reveal whether indeed such trade-off between bacterial motility and root adherence
341 underlies our observations in this evolutionary experiment.

342 The decreased production of public goods, such as siderophores and exoprotease, which were observed
343 in all independent evolutionary lines (Figure 5), could be beneficial for bacterial fitness by saving energy
344 and primary metabolites. For example, adaptation of *Pseudomonas aeruginosa* to the human host
345 through mutations in regulators was accompanied by loss of siderophore production, secreted proteases,
346 and biofilm formation (66, 67).

347

348 **Cell surface structures as bacterial adaptation targets**

349 Bacterial surface components are the first line of defense against environmental stress and interplay with
350 hosts (68–70). Lipopolysaccharide (LPS) is a central outer membrane component for gram-negative
351 bacteria, and exhibits structural adaptability that is contributed especially by its O-antigen part (68, 70).
352 Bacterial O-antigen structure modification plays an important role in evasion of host immunity (71), and
353 has the potential to change host-bacteria interactions (68, 70, 72). In plant pathogenic bacteria LPS
354 components are important virulence determinants (73, 74), that activate a variety of defense-related
355 responses (73, 75–77). In *P. fluorescens* the O-antigen component has been implicated to induce systemic
356 resistance in radish (78) and Arabidopsis (79).

357 We observed parallel mutations in genes that are involved in LPS biosynthesis and structure modification.
358 In three out of the five evolutionary lines, the first fixed mutations were identified in *oafA* and *galE* that
359 are annotated as O-antigen biosynthesis and structure modification related (Table 1). *OafA* encodes an O-
360 acetyltransferase, which is postulated to modify the O-antigen by acetylation (32). The enzyme GalE, UDP-
361 galactose 4-epimerase, is involved in the interconversion of UDP-glucose to UDP-galactose, an essential
362 intermediate for LPS core and O-antigen structures (33–35). Inactivation of *galE* in *Porphyromonas*
363 *gingivalis* resulted in shortening of the O-antigenic side chain (80). And in *Bradyrhizobium japonicum*
364 disruption of *galE* resulted in complete lack of the O-antigen (33). Thus, it is tempting to speculate that
365 evasion of the plant's immune response plays a role in adaptation of CHA0 to the rhizosphere of
366 Arabidopsis.

367 In conclusion, the observed bacterial genetic and phenotypic adaption dynamics emphasize an important
368 role of global regulators, motility, and cell surface structures in bacterial adaptation to its host. The parallel

369 emergence of mutations in similar genes resulting in specific fitness advantages in the rhizosphere,
370 suggest that this evolutionary process is driven by the rhizosphere environment.

371

372 ## Materials and Methods

373

374 ### Experimental setup

375 We set up an experimental evolution experiment with *Arabidopsis thaliana* (Arabidopsis) as host plant
376 and *Pseudomonas protegens* CHA0 (CHA0) as the evolving bacterial strain. CHA0 (81) is a model strain
377 originally isolated from roots of tobacco plants grown in soil naturally suppressive to black root rot (82).
378 CHA0 was chromosomally tagged with GFP and a kanamycin resistance cassette (Jousset et al 2006) to
379 enable easy tracking of the strain and checking for contaminations. Briefly, the ancestral bacterial
380 population was inoculated on Arabidopsis roots in a gnotobiotic system. After a plant growth cycle of 4
381 weeks, evolved bacterial populations were transferred to new plants and this cycle was repeated 8 times.
382 Bacterial colonies were retrieved from the evolved populations at the end of each growth cycles by
383 plating.

384 We selected six bacterial isolates from each independent evolutionary line at the end of plant growth
385 cycles two, four and six. Together with six isolates from the ancestor population, for a total of 94 isolates
386 various life-history traits (as shown below) were measured and their genomes re-sequenced using the
387 NextSeq-500 Illumina platform (2 x 75 bp paired-end). We used the snippy pipeline
388 (https://github.com/tseemann/snippy), integrating reference genome-based mapping of the Illumina
389 reads by BWA-MEM, variant calling by SAMtools and FreeBayes, and variant impact analysis using SnpEff
390 (83), to identify single nucleotide polymorphisms (SNPs) and small indels and deletions (INDELs). Larger
391 INDELs were identified by calculating the breath of coverage of the mapped Illumina reads on the
392 reference genome in a sliding window using bedtools (84). Regions with reduced coverage (<99%) were
393 manually inspected in the Integrative Genome Viewer (IGV). For two isolates from experimental line four
394 re-sequencing failed.

395

396 ### Bacterial life-history traits

397 A variety of bacterial life-history traits, reflecting various aspects of bacterial physiological processes, were
398 measured. We measured bacterial growth yield and resistance to various stresses, including acidic (pH =

399 5) and alkaline (pH = 9) conditions, oxidative stress in 0.0025% H_2O_2, water potential stress (15%
400 polyethylene glycol (PEG)-6000)), and salt stress (2% NaCl), and resistance to the antibiotics streptomycin
401 (1 µg · ml^{-1}), tetracycline (1 µg · ml^{-1}), and penicillin (5 µg · ml^{-1}). Bacterial growth yield was quantified in
402 modified Ornston and Stanier (OS) minimal medium (85) supplemented with single carbon sources that
403 have been reported to be abundant in Arabidopsis root exudates (Chaparro et al 2013). These included
404 the following carbon sources; alanine, arabinose, butyrolactam, fructose, galactose, glucose, glycerol,
405 glycine, lactic acid, putrescine, serine, succinic acid, threonine, and valine that were added to a final
406 concentration of 0.5 g · l^{-1}. Growth yield was measured as OD_{600} (SPECTROstar Nano plate reader), and
407 this measure was used to determine carbon source utilization ability. In addition, we measured bacterial
408 auxin (indole-3-acetic acid or IAA) production with a colorimetric test (86), iron-chelating ability using a
409 modified chrome azurol S (CAS) assay (87), proteolytic activity by the adapted assay from Smeltzer et al.
410 (1993), tryptophan side chain oxidase activity using a colorimetric assay (88), and biofilm formation using
411 a modified Chrystal Violet staining assay (89). We further assessed bacterial antimicrobial activity by
412 quantifying their effect on growth of the fungi *Verticillium dahliae* and *Fusarium oxysporum,* and the
413 bacterium *Ralstonia solanacearum*. Bacterial motility assays were performed on Cook's Cytophaga (CC)
414 medium containing 0.2% tryptone (90). This medium was solidified with 0.3 or 0.5% agar to determine
415 swimming and swarming motility, respectively (42). Bacteria were first cultured in liquid CC medium
416 overnight at 28 °C, and subsequently inoculated in the middle of the motility plates using toothpicks.
417 Plates were incubated upside-down at 28 °C for 24 h, and the colony areas were calculated using Image J
418 as a measure of bacterial motility.

419

420 **Hierarchical and Model-based clustering of bacterial traits**

421 *Hierarchical clustering*

422 A heatmap to illustrate the association patterns of bacterial genotypes and their measured traits was
423 constructed in R using the *ggplot2* package. Hierarchical clustering was performed using the Ward.D2
424 method, that is based on the squared dissimilarities of the set of objects being clustered (91).

425 *Model-based clustering*

426 We applied a model-based clustering method to reveal the best fitting structures of trait covariance
427 patterns. For example, some traits might be either directly or indirectly co-regulated by the same gene,
428 which is expected for global regulators particularly which can co-regulate thousands of genes. We used

429 the *mclust* package in R to run the model simulation (92). This method assumes that the input data are
430 distributed as a mixture of two or more clusters. The advantage of the model-based clustering method is
431 that it avoids heuristic assumptions and uses a soft assignment that every data point has a possibility of
432 falling to each cluster, which facilitates the best clustering solution. The so-called Bayesian Information
433 Criterion (BIC) was used to select the best model. A large BIC score indicates a better fit of the model.

434 This result is in line with the outcome of hierarchical clustering with Adjusted Rand Index (ARI) set as 1,
435 and *k* set as 4 in "ward.D2" as indicated in Figure 5a. ARI is usually used to evaluate the match degree of
436 a given clustering solution comparing to the model-based clustering result, with 0 reflecting a random
437 partition and 1 the boundary of accuracy of a certain clustering solution (92).

438

439 **Genotype-phenotype association analysis**

440 Bacterial traits within each model-predicted cluster have similar data distribution patterns and covaried
441 together by the definition of the clustering method. Thus, we applied a linear regression-based method,
442 i.e. principle component analysis (PCA), to reduce the dimensionality of data and generate a proxy for
443 each model predicted cluster. These proxies were later used as the x-axis values in Figure 5b-e. We applied
444 the package *ggbiplot* in R to generate the PCA plots and PC1 index from the normalized datasets. The
445 proxies were normalized for further analysis.

446 To examine the effect of each mutation on bacterial phenotype, a Students' *t*-test was used to compare
447 cluster proxies of evolved genotypes with their direct ancestors. Only genotypes identified more than
448 once were included in this analysis.

449

450 **Relative quantification of mutant frequency using HRM profile analysis**

451 We used High-Resolution Melting (HRM) profile analysis with integrated LunaProbes to quantify the ratio
452 of mutant to wild type genotypes (93–95). The probes and primers used in this study are listed in Table
453 S2. Primers were designed using Primer3. Probes were designed with the single nucleotide polymorphism
454 (SNP) located in the middle of the sequence, and the 3' end was blocked by carbon spacer C3. The primer
455 asymmetry was set to 2:1 (excess primer: limiting primer) in all cases. Pre-PCR was performed in a 10-µl
456 reaction system, with 0.25 µM excess primer, 0.125 µM limiting primer, 0.25 µM probe, 0.5 µl bacterial
457 sample culture (100-fold diluted saved sample, OD_{600} is about 0.01), 1X LightScanner Master Mix (BioFire

458 Defense). DMSO with the final concentration 5% was supplemented in all reactions to ensure the targeted
459 melting domains are within the detection limit of the LightScanner (Idaho Technology Inc.). Finally, MQ
460 water was used to supplement up to 10 µl. A 96-well black microtiter plate with white wells was used to
461 minimize background fluorescence. Before amplification, 25 µl mineral oil was loaded in each well to
462 prevent evaporation, and the plate was covered with a foil seal to prevent the degradation of fluorescent
463 molecules. Amplification was initiated by a holding at 95 °C for 3 min, followed by 55 cycles of
464 denaturation at 95 °C for 30 s, annealing at 60 °C for 30 s and extension at 72 °C for 30 s and then kept at
465 72 °C for 10 min. After amplification, samples were heated in a ThermalCycler (Bio-Rad) shortly to 95 °C
466 for 30 s to denature all double-stranded structures followed by a rapid cooling to 25 °C for 30 s to facilitate
467 successful hybridization between probes and the target strands. The plate was then transferred to a
468 LightScanner (Idaho Technology Inc.). Melting profiles of each well were collected by monitoring the
469 continuous loss of fluorescence with a steady increase of the temperature from 35 °C to 97 °C with a ramp
470 rate of 0.1 °C /s. The relative quantification was based on the negative first derivative plots using software
471 MATLAB. The areas of probe-target duplexes melting peaks were auto-calculated by 'AutoFit Peaks I
472 Residuals' function in software PeakFit (SeaSolve Software Inc.). The mutant frequency *X* was calculated
473 using the formula shown below:

474 $$X = \frac{\text{Area}_{mutant}}{\text{Area}_{mutant} + \text{Area}_{WT}}$$

475 To validate the HRM method, standard curves were generated by measuring mixed samples with known
476 proportions of mutant templates: 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%.
477 Measurements for each sample were done in triplicate. Linear regression formula of each mutant
478 between actual frequencies and measured frequencies were shown in Figure S1. The high R^2 values, and
479 nearly equal to 1 slope values of these equations, confirmed that the HRM method can accurately detect
480 mutants' frequency in a mixed population.

481

482 **Data availability**

483 The *P. protegens* CHA0-GFP reference genome is deposited on GenBank: RCSR00000000.1. Raw
484 sequencing data used in this study are deposited at the NCBI database under BioProject PRJNA473919. A
485 conversion table for the CHA0-GFP to CHA0 gene annotations including recent NCBI accession codes is
486 available here: https://doi.org/10.6084/m9.figshare.13295828.v2.

487

Acknowledgments

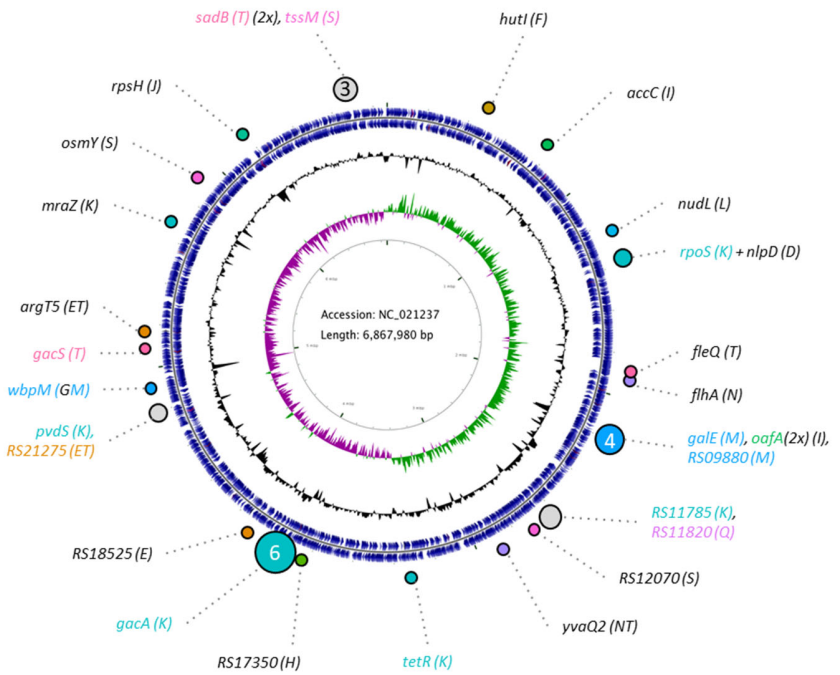
This work was supported by a China Scholarship Council fellowship (to E.L.), a postdoctoral fellowship of the Research Foundation Flanders (FWO 12B8116RN) (to R.D.J.), and a European Research Council Advanced Grant 269072 (to C.M.J.P.).

Author contributions

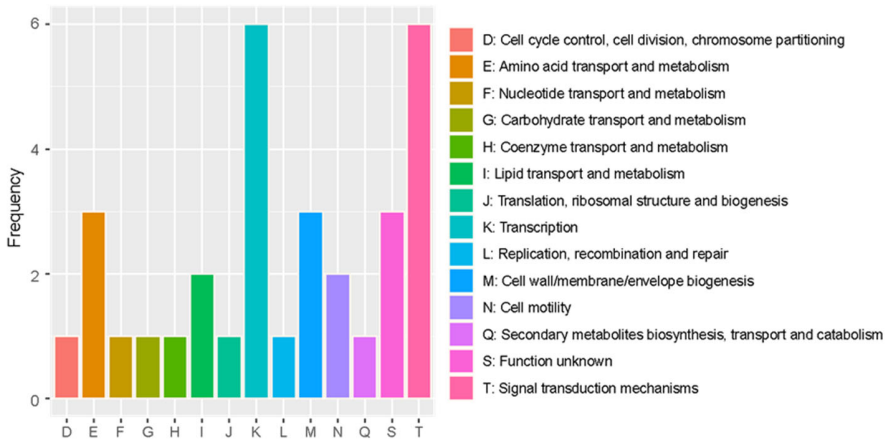
E.L., A.J., C.M.J.P., P.A.H.M.B., and R.D.J. designed experiments; E.L. and H.J. performed experiments; E.L. and R.D.J. analyzed data. All authors discussed the results and contributed to the final manuscript.

497 **Figures and Figure legends**

a)



b)



498

499 **Figure 1.** Genomic distribution of *Pseudomonas protegens* CHA0 evolutionary adaptations. **a)** Rings from inside to outside: Ring
500 1, nucleotide position indicator; Ring 2, green/purple GC-skew (-/+); Ring 3, %GC; Ring 4, protein-coding genes; Ring 5, distribution
501 of identified mutations arisen during the evolutionary experiment. Functional annotation of mutated genes are indicated by color
502 representing Cluster of orthologous groups (COGs); the key of which can be seen in panel B. **b)** Frequency of mutations per COG
503 class, highlighting enrichment in the classes T (signal transduction), K (transcription), M (cell wall) and E (amino acid transport
504 and metabolism).

505

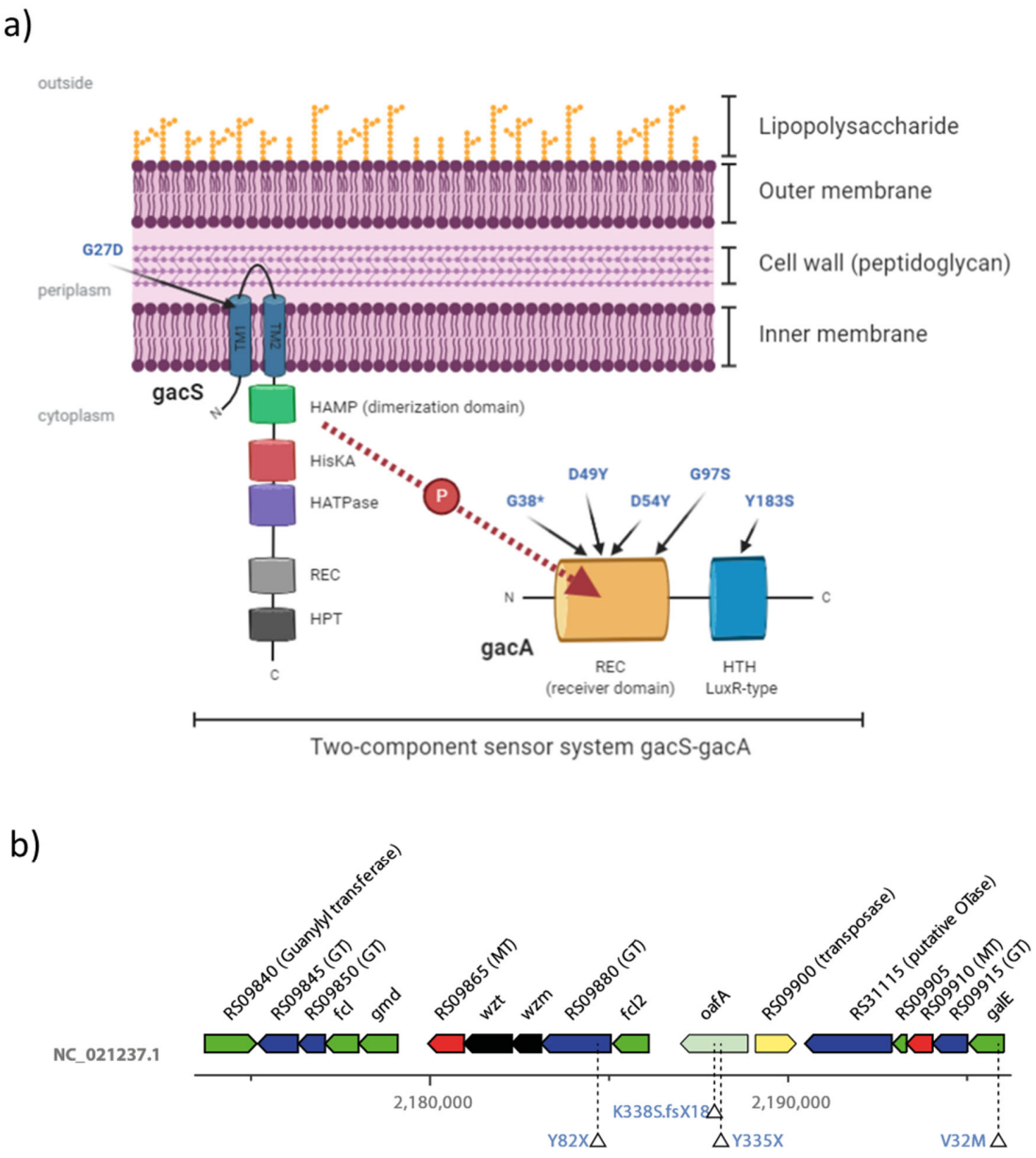


Figure 2. Localization of *gac* and *OBC3* mutations identified in the evolutionary experiment. **a)** Schematic view of the cell wall of the gram-negative bacterium *Pseudomonas protegens* highlighting the presence of the lipopolysaccharide (LPS) on the outside and the presence of the two-component regulator system GacS/GacA on the inner membrane and in the cytoplasm. The arrows indicate the location of the amino acid substitutions that were identified in this study in GacS (G27D) and GacA (G38X, D49Y, D54Y, G97S and Y183S). Note the travel of a phosphate group (P) from GacS to GacA upon signal perception which is accepted by either the Asp49 or Asp54 residue in GacA which are both found mutated in two different mutants. **b)** Genomic region harboring the *OBC3* gene cluster responsible for the synthesis of long O-polysaccharide (O-PS) on the LPS and indicated mutations that were identified in a glycosyl transferase (RS09880; Y82X), the O-antigen acetyltransferase oafA (K338S.fsX18 and Y335X) and the UDP-glucose 4-epimerase (galE; V32M).

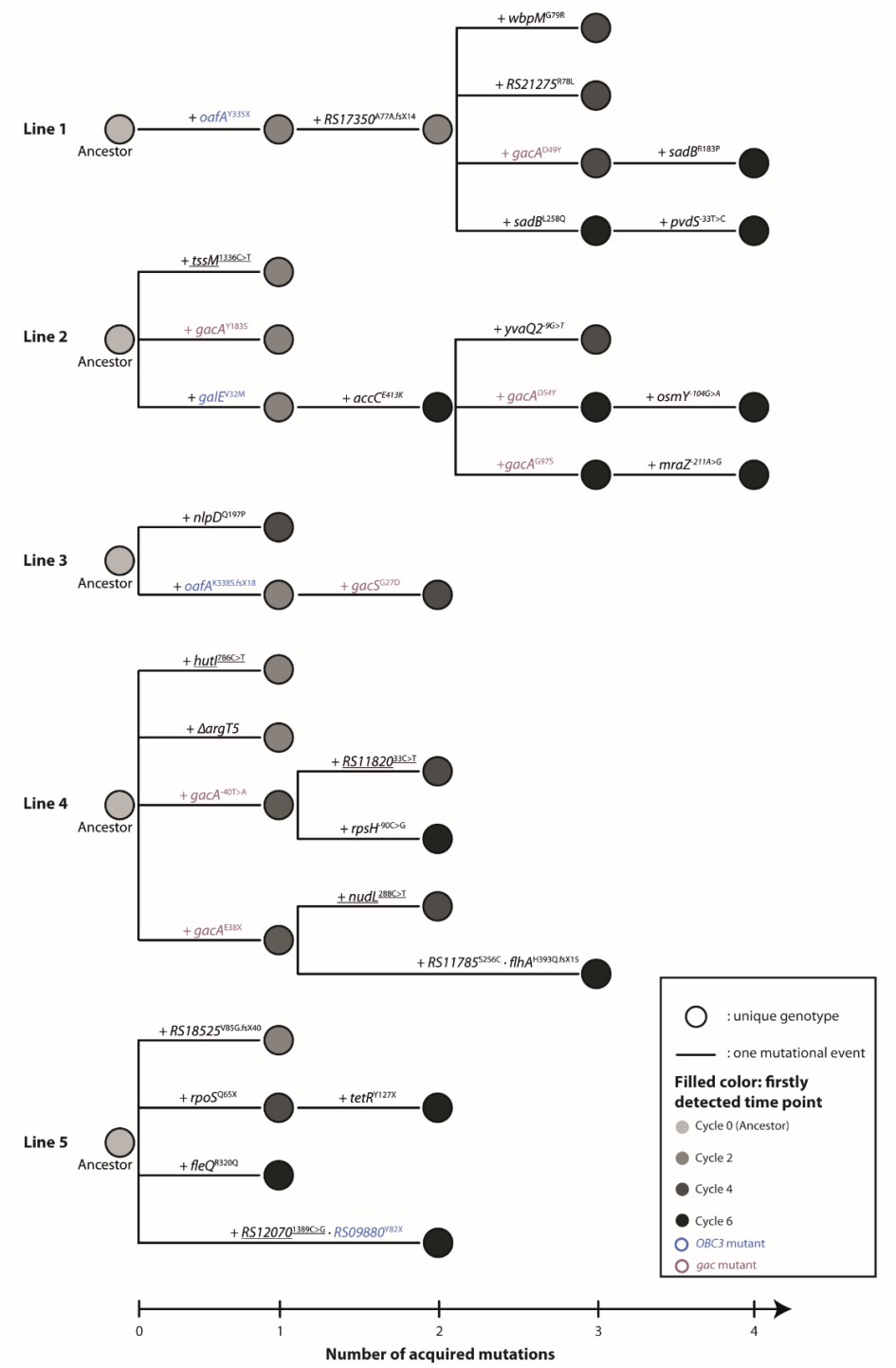


Figure 3. Phylogenetic trees of five independently evolving *P. protegens* CHA0 populations. Phylogenies for 18 genomes from each population (16 for the population in experimental line 4), based on the sequential appearance of mutations. Synonymous mutations are underlined. Circles represent unique bacterial genotypes. Intensity of the circle fill represents the time point at which a genotype was firstly detected. *OBC3* and *gac* mutants specifically are highlighted in blue and purple respectively. The frequency of each mutation in the respective population can be found in Table 2.

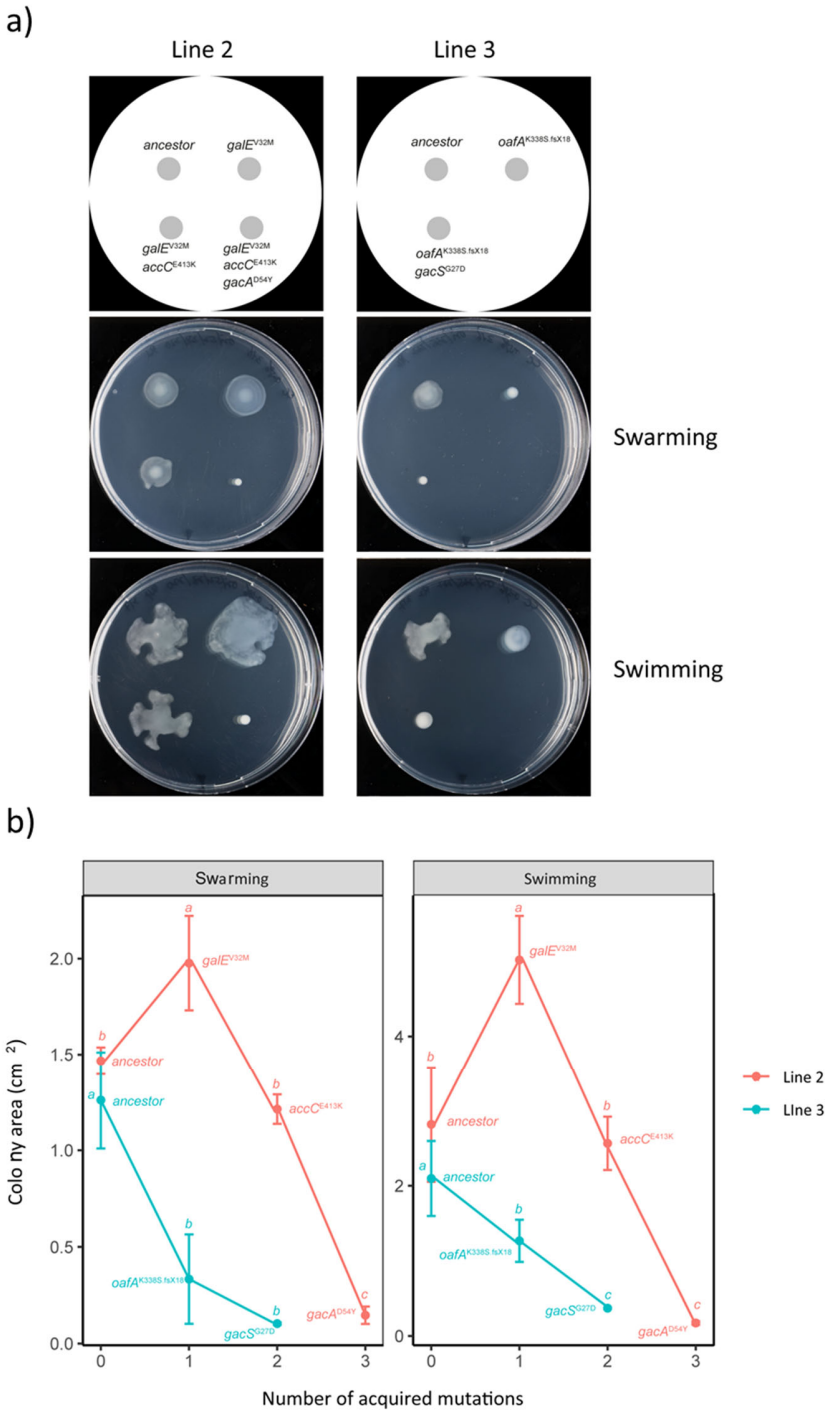
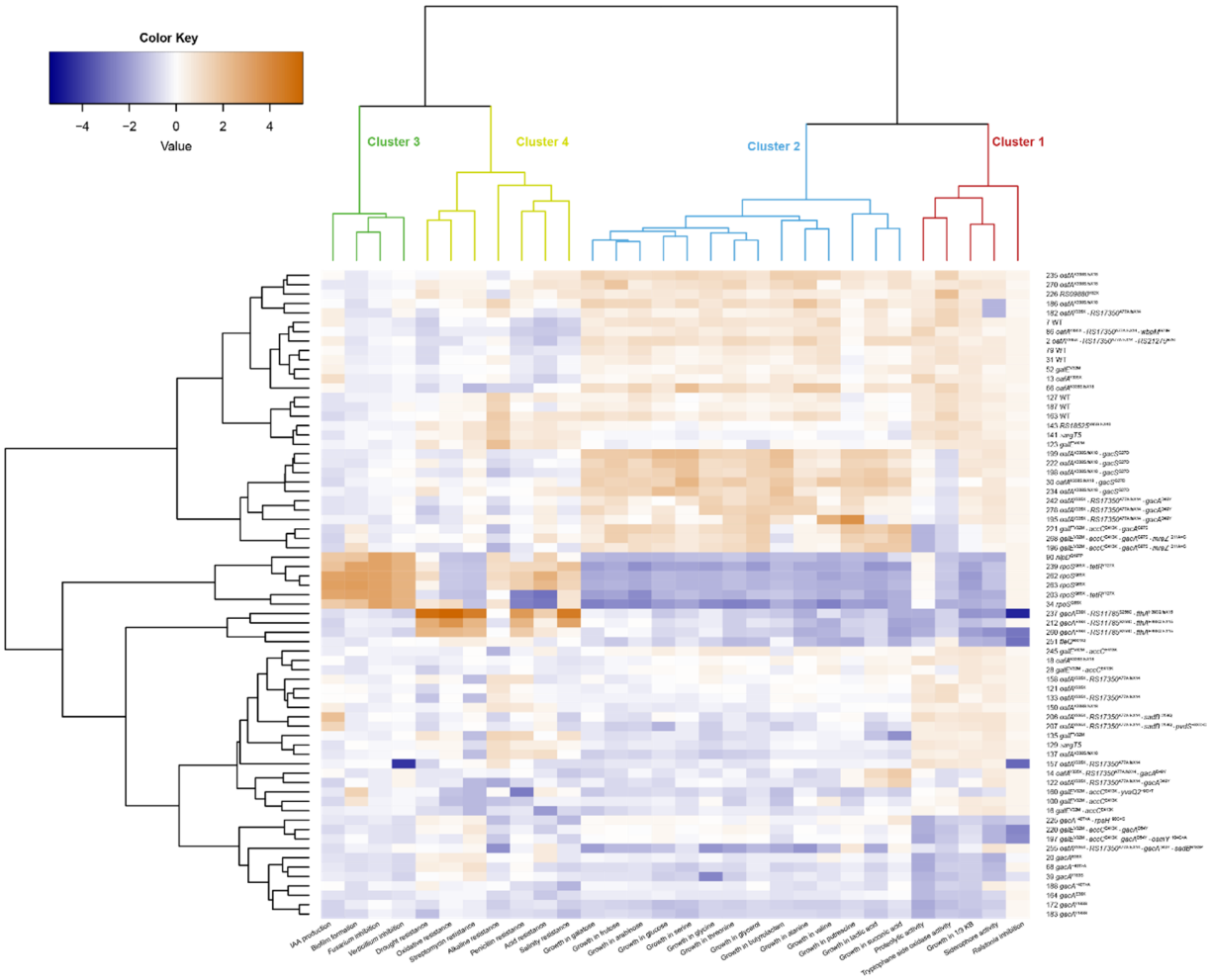


Figure 4. Effects of *gacS/gacA* mutations on bacterial swimming and swarming behaviour. **a)** The top panels illustrate the experimental layout of motility assays, that are pictured in the middle and bottom panels for swarming and swimming respectively. **b)** The accumulative effects of mutations on swarming (left panel) and swimming (right panel) behaviour. Colony area as a measure for bacterial motility was determined using ImageJ. The number of acquired mutations compared to the ancestor is used to express the genetic distance.

a)


Color Key
−4
−2
0
2
4
Value
Cluster 3
Cluster 4
Cluster 2
Cluster 1
IAA production
Biofilm formation
Furanium inhibition
Verticillium inhibition
Drought resistance
Oxidative resistance
Streptomycin resistance
Alkaline resistance
Penicillin resistance
Acid resistance
Salinity resistance
Growth in galactose
Growth in frutose
Growth in arabinose
Growth in glucose
Growth in serine
Growth in glycine
Growth in threonine
Growth in glycerol
Growth in butyrolactam
Growth in alanine
Growth in valine
Growth in pathenine
Growth in lactic acid
Growth in succinic acid
Proteolytic activity
Tryptophane side oxidase activity
Growth in 1/3 KB
Siderophore activity
Rafatoxia inhibition

528

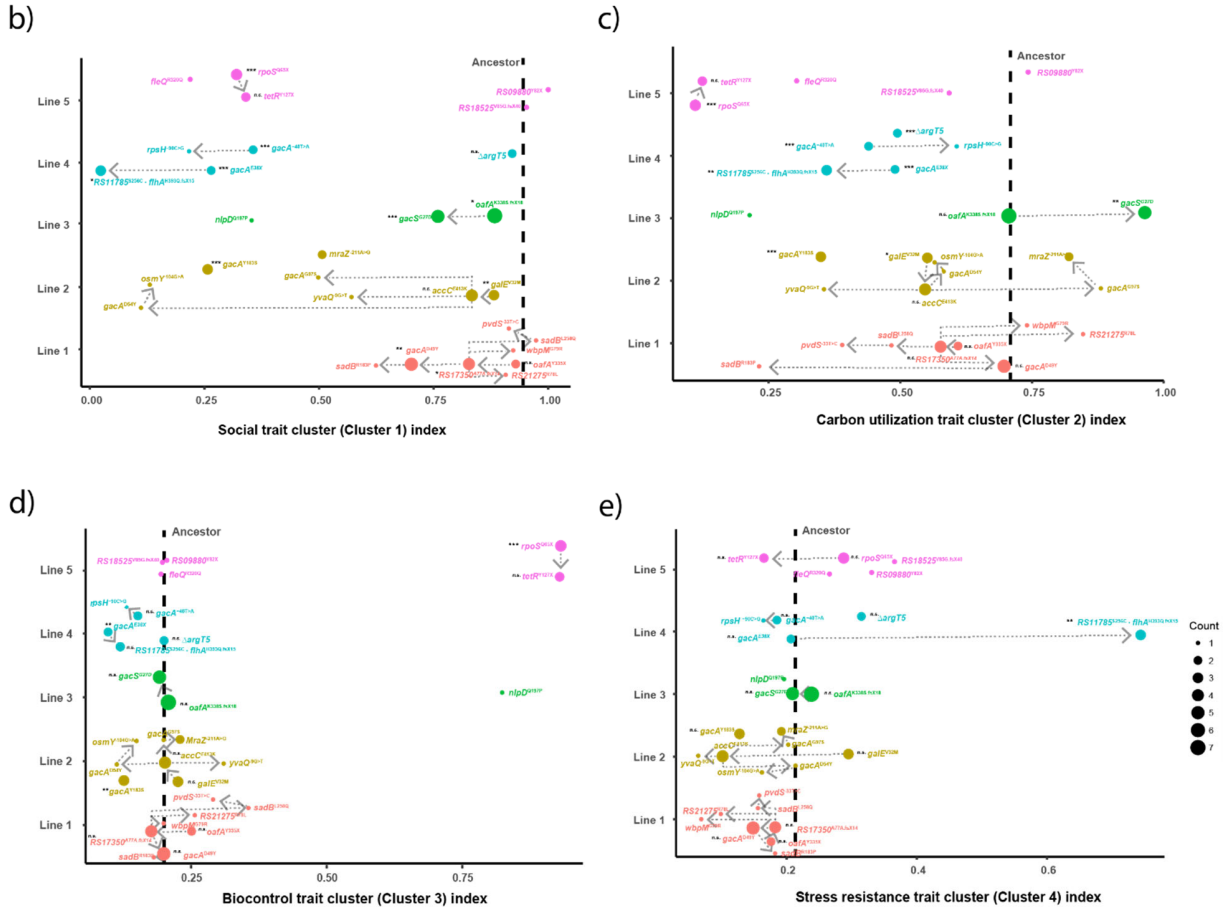
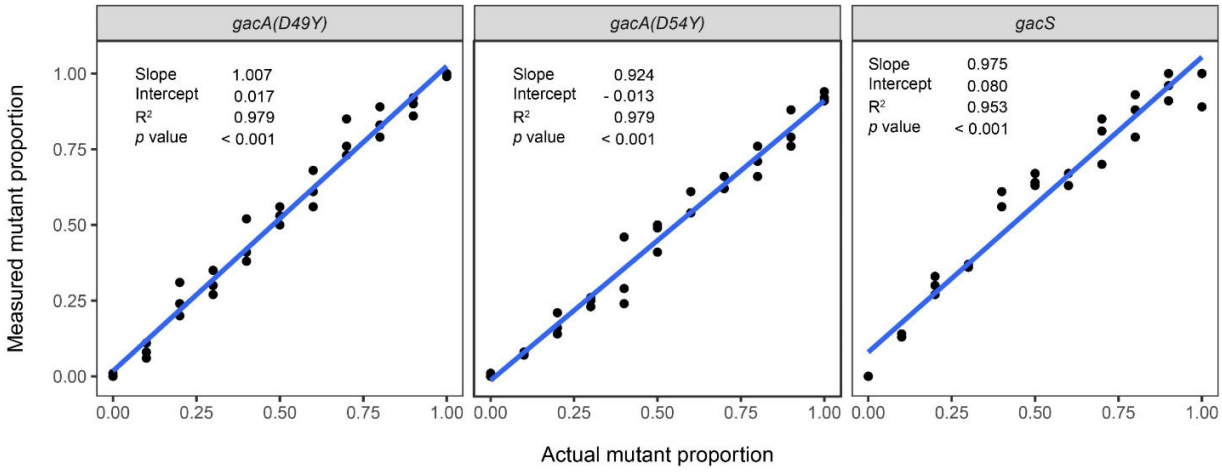


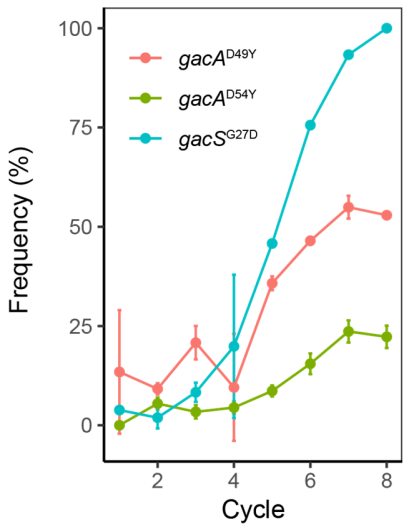
Figure 5. Genotype-phenotype association analysis of evolved bacterial populations. **a)** Heatmap representing the performance of 30 life-history traits of six ancestral and 63 evolved isolates (isolates without mutations or only with synonymous mutation were excluded for this association analysis). The "ward.D2" method was used for hierarchical clustering, based on the dissimilarities of the set of objects being clustered. The four clusters depicted on top are model-predicted. Each column illustrates one bacterial life-history trait, indicated on the bottom of the figure. Each row represents an isolate, indicated by its sample ID number and mutational events. **b-e)** Principle component analysis was applied to generate a general symbolic index for each model predicted cluster. The proxy values of each cluster were normalized, higher values represent better performance. To illustrate the effects of each mutational event, ANOVA was applied to reveal the phenotypic change between each genotype and its identified nearest ancestor in each line. Colour-filled circles represent bacterial genotypes, with different colours representing independent evolving populations and the size representing the count. The dashed line represents the average value of the ancestor.

541 **Supplementary Figures:**



542

543 **Figure S1.** Standard curves of measured mutant versus ancestor proportion as a function of the actual proportion,
544 using series of mixed samples with known proportions (0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and
545 100% of mutant frequency). Relative densities of mutants *gacA*^{D49Y}, *gacA*^{D54Y}, and *gacS*^{G27D} were measured by PCR-
546 based high-resolution melting (HRM) analysis. Measurements for each sample were performed in triplicate. In each
547 plot, the black dots represent the measurements, the blue line the fit which was generated based on linear
548 regression modelling.



549

550 **Figure S2.** Frequency trajectories of *gacA*^{D49Y}, *gacA*^{D54Y} and *gacS*^{G27D} mutants during long-term rhizosphere adaption.
551 The X-axis represents the plant-to-plant transferring cycle of the bacterial population. Mutant frequency was
552 determined by quantifying the ratio of mutant allele relative to wild-type allele, using PCR-based high-resolution
553 melting (HRM) analysis. The data shown are the mean of two technical replicates, and error bars represent the
554 standard deviation of the mean.

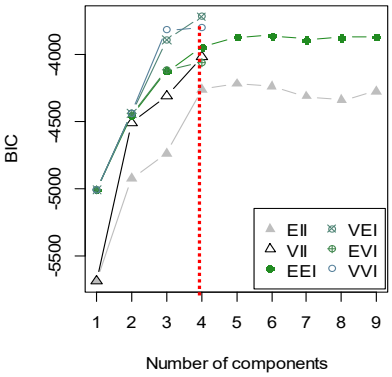
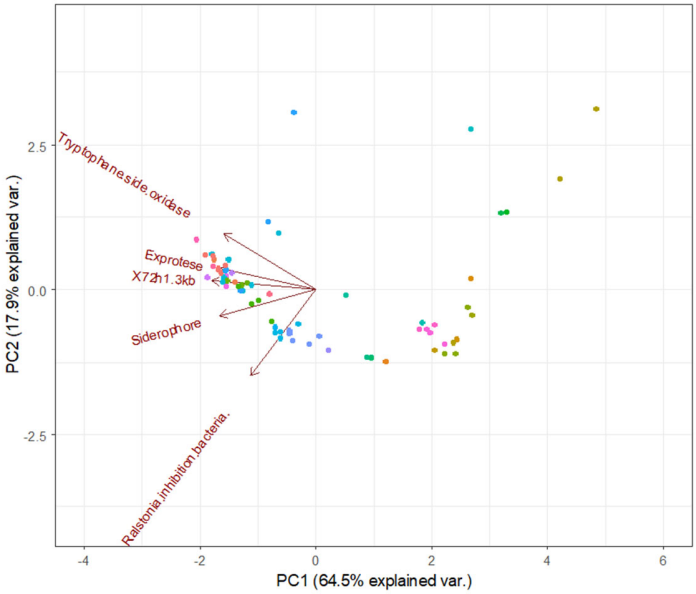


Figure S3. Selection of the best clustering model using Bayesian Information Criterion (BIC) using the *mclust* R package. A large BIC score indicates strong evidence for the corresponding model. The VEI model with 4 components best fits our data. VEI indicates that the clusters have variable volume, the same shape and an orientation equal to coordinate axes.

a



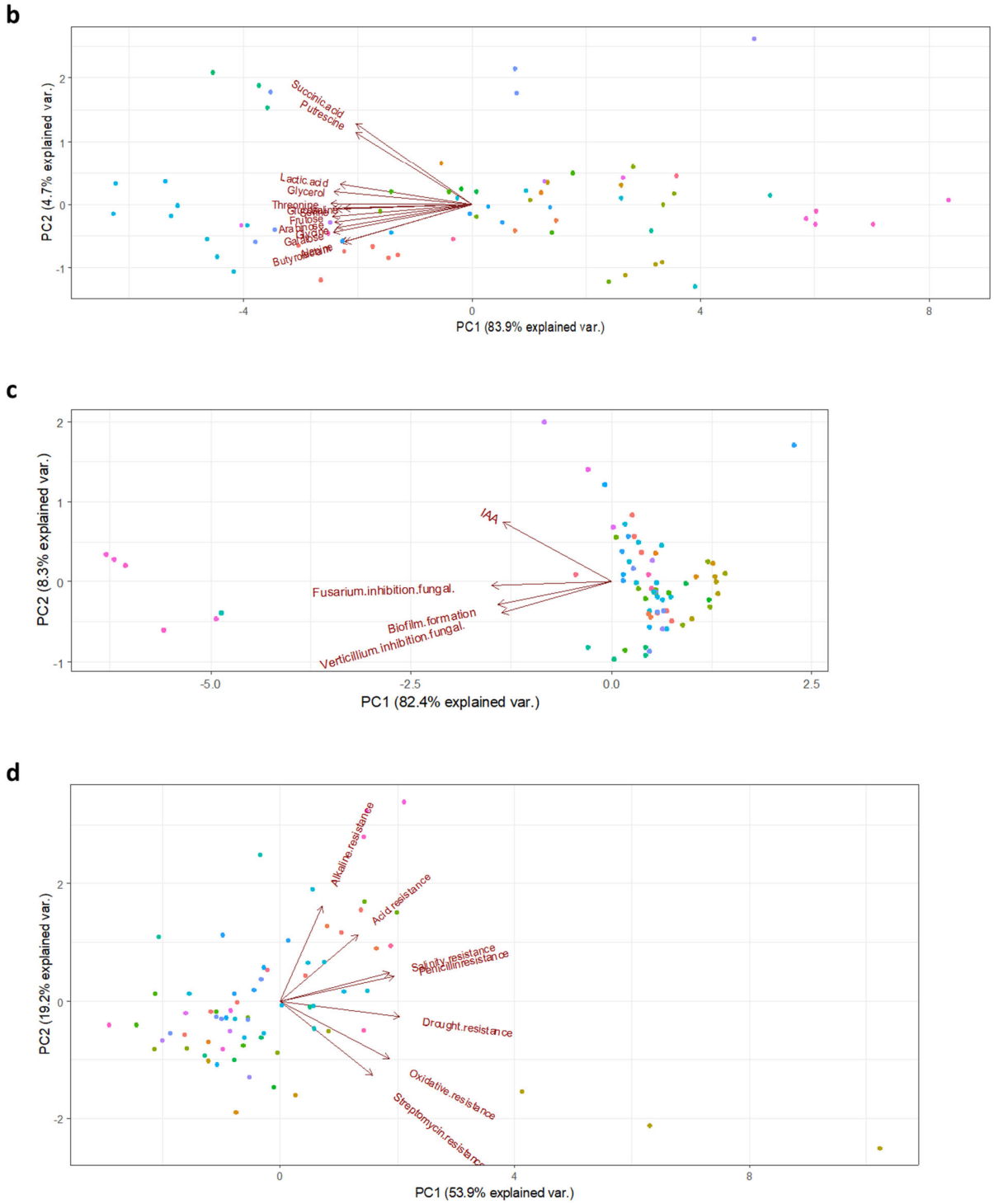


Figure S4. Coordinate axes transformation of each model predicted cluster using principal component analysis (PCA). **a.** PCA plot of social trait cluster (cluster 1); **b.** PCA plot of carbon utilization trait cluster (cluster 2); **c.** PCA plot of

566 biocontrol trait cluster (cluster 3); **d.** PCA plot of stress resistance trait cluster (cluster 4). Each dot represents one
567 isolate.

568 **Table S1** Genotypes of whole-genome sequenced CHA0 isolates.

Sample ID	Line	Cycle	Genotype
61	1	2	*WT*
169	1	2	*WT*
13	1	2	*oafA*^{Y335X}
121	1	2	*oafA*^{Y335X}
133	1	2	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$
157	1	2	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$
158	1	4	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$
182	1	4	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$
14	1	4	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$ · *gacA*^{D49Y}
122	1	4	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$ · *gacA*^{D49Y}
86	1	4	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$ · *wbpM*^{G79R}
2	1	4	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$ · *RS21275*^{R78L}
195	1	6	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$ · *gacA*^{D49Y}
242	1	6	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$ · *gacA*^{D49Y}
278	1	6	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$ · *gacA*^{D49Y}
255	1	6	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$ · *gacA*^{D49Y} · *sadB*^{R183P}
206	1	6	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$ · *sadB*^{L258Q}
207	1	6	*oafA*^{Y335X} · *RS17350*$^{A77A.fsX14}$ · *sadB*^{L258Q} · *pvdS*$^{-33T>C}$
87	2	2	*WT*
39	2	2	*gacA*^{Y183S}
183	2	2	*gacA*^{Y183S}
123	2	2	*galE*^{V32M}
135	2	2	*galE*^{V32M}
63	2	2	*tssM*$^{1336C>T}$
172	2	4	*gacA*^{Y183S}
52	2	4	*galE*^{V32M}
16	2	4	*galE*^{V32M} · *accC*^{E413K}
28	2	4	*galE*^{V32M} · *accC*^{E413K}
100	2	4	*galE*^{V32M} · *accC*^{E413K}
160	2	4	*galE*^{V32M} · *accC*^{E413K} · *yvaQ2*$^{-9G>T}$
245	2	6	*galE*^{V32M} · *accC*^{E413K}
220	2	6	*galE*^{V32M} · *accC*^{E413K} · *gacA*^{D54Y}

197	2	6	*galE*^{V32M} · *accC*^{E413K} · *gacA*^{D54Y} · *osmY*$^{-104G>A}$
221	2	6	*galE*^{V32M} · *accC*^{E413K} · *gacA*^{G97S}
196	2	6	*galE*^{V32M} · *accC*^{E413K} · *gacA*^{G97S} · *mraZ*$^{-211A>G}$
268	2	6	*galE*^{V32M} · *accC*^{E413K} · *gacA*^{G97S} · *mraZ*$^{-211A>G}$
29	3	2	*WT*
65	3	2	*WT*
53	3	2	*WT*
77	3	2	*WT*
161	3	2	*WT*
137	3	2	*oafA*$^{K338S.fsX18}$
90	3	4	*nlpD*^{Q197P}
18	3	4	*oafA*$^{K338S.fsX18}$
66	3	4	*oafA*$^{K338S.fsX18}$
150	3	4	*oafA*$^{K338S.fsX18}$
186	3	4	*oafA*$^{K338S.fsX18}$
30	3	4	*oafA*$^{K338S.fsX18}$ · *gacS*^{G27D}
235	3	6	*oafA*$^{K338S.fsX18}$
270	3	6	*oafA*$^{K338S.fsX18}$
199	3	6	*oafA*$^{K338S.fsX18}$ · *gacS*^{G27D}
222	3	6	*oafA*$^{K338S.fsX18}$ · *gacS*^{G27D}
234	3	6	*oafA*$^{K338S.fsX18}$ · *gacS*^{G27D}
198	3	6	*oafA*$^{K338S.fsX18}$ · *gacS*^{G27D}
57	4	2	*WT*
105	4	2	*WT*
81	4	2	*WT*
129	4	2	*ΔargT5*
141	4	2	*ΔargT5*
21	4	2	*hult*$^{786C>T}$
80	4	4	*WT*
152	4	4	*WT*
188	4	4	*gacA*$^{-40T>A}$
68	4	4	*gacA*$^{-40T>A}$ · *RS11820*$^{33C>T}$
164	4	4	*gacA*^{E38X}
20	4	4	*gacA*^{E38X} · *nudL*$^{288C>T}$
225	4	6	*gacA*$^{-40T>A}$ · *rpsH*$^{-90C>G}$
212	4	6	*gacA*^{E38X} · *RS11785*^{S256C} · *flhA*$^{H393Q.fsX15}$

237	4	6	*gacA*^{E38X} · *RS11785*^{S256C} · *flhA*$^{H393Q.fsX15}$
260	4	6	*gacA*^{E38X} · *RS11785*^{S256C} · *flhA*$^{H393Q.fsX15}$
11	6	2	*WT*
71	6	2	*WT*
191	6	2	*WT*
35	6	2	*WT*
83	6	2	*WT*
143	6	2	*RS18525*$^{V85G.fsX40}$
70	6	4	*WT*
106	6	4	*WT*
142	6	4	*WT*
58	6	4	*WT*
190	6	4	*WT*
34	6	4	*rpoS*^{Q65X}
251	6	6	*fleQ*^{R320Q}
262	6	6	*rpoS*^{Q65X}
263	6	6	*rpoS*^{Q65X}
203	6	6	*rpoS*^{Q65X} · *tetR*^{Y127X}
239	6	6	*rpoS*^{Q65X} · *tetR*^{Y127X}
226	6	6	*RS09880*^{Y82X} · *RS12070*$^{1389C>G}$
7	ancestor	0	*WT*
31	ancestor	0	*WT*
79	ancestor	0	*WT*
127	ancestor	0	*WT*
163	ancestor	0	*WT*
187	ancestor	0	*WT*

569 Note: X, represents a stop codon (at its relative position in case of a shifted frame); fs, frame shift; del, deletion; DNA sequence
570 change positions are relative to the cDNA.

571

572 **Table S2** Primers and probes used for high-resolution melting (HRM) analysis. For the two *gacA* mutants the same set of primers
573 was used. Underlined bases indicate the position of the single nucleotide point (SNP) mutations within the probe sequences.
574 ΔTm (°C) indicates the melting temperature difference between WT-probe duplex and mutant-probe duplex.

Target gene	Strain ID	SNP locus	Forward primer (excess)	Reverse primer (limiting)	Amplicon size	Probe sequence	Probe length	Target strand	Perfect match/ mismatch	ΔTm (°C)
gacA^{D49Y}	242	145G>T	5'-ATCGATGGCCTGCAAGTAGT-3'	5'-CGGGTAGGAAAGGGATCTTC-3'	206 bp	5'-CATCAGGACCACATCGGGCTTCAGCTCCCG-/C3/3'	30nt	WT	G::C / T::C	5.31
gacA^{D54Y}	220	160G>T				5'-TGGCATCTTGACGTCCATCAGGACCACATC-/C3/3'	30nt	WT	G::C / T::C	4.92
gacS^{G27D}	222	80G>A	5'-GCGTACTGTTGCTGACCTTG-3'	5'-AGCATCTGGGTGTTGTGGTT-3'	178bp	5'-AGGTGAAGTAGCCGCCCAGCACCAAAGCCA-/C3/3'	30nt	WT	G::C / A::C	4.83

575

576 **References**

577 1. Berendsen RL, Pieterse CMJ, Bakker PAHM. 2012. The rhizosphere microbiome and plant health. Trends
578 Plant Sci 17:478–486.
579 2. Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM. 2006. The role of root exudates in rhizosphere interations
580 with plants and other organisms. Annu Rev Plant Biol 57:233–266.
581 3. Stringlis IA, Yu K, Feussner K, de Jonge R, Van Bentum S, Van Verk MC, Berendsen RL, Bakker PAHM,
582 Feussner I, Pieterse CMJ. 2018. MYB72-dependent coumarin exudation shapes root microbiome assembly
583 to promote plant health. Proc Natl Acad Sci U S A 115:E5213–E5222.
584 4. Bakker PAHM, Pieterse CMJ, de Jonge R, Berendsen RL. 2018. The soil-borne legacy. Cell 172:1178–1180.
585 5. Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH. 2013. Going back to the roots: the microbial
586 ecology of the rhizosphere. Nat Rev Microbiol 11:789–799.
587 6. Sasse J, Martinoia E, Northen T. 2018. Feed your friends: do plant exudates shape the root microbiome?
588 Trends Plant Sci 23:25–41.
589 7. Bais HP, Park SW, Weir TL, Callaway RM, Vivanco JM. 2004. How plants communicate using the
590 underground information superhighway. Trends Plant Sci 9:26–32.
591 8. Zamioudis C, Pieterse CMJ. 2012. Modulation of host immunity by beneficial microbes. Mol Plant-Microbe
592 Interact 25:139–150.
593 9. Yang J, Kloepper JW, Ryu C-M. 2009. Rhizosphere bacteria help plants tolerate abiotic stress. Trends Plant
594 Sci 14:1–4.
595 10. Browne P, Rice O, Miller SH, Burke J, Dowling DN, Morrissey JP, O'Gara F. 2009. Superior inorganic
596 phosphate solubilization is linked to phylogeny within the *Pseudomonas fluorescens* complex. Appl Soil Ecol
597 43:131–138.
598 11. Lugtenberg B, Kamilova F. 2009. Plant growth promoting rhizobacteria. Annu Rev Microbiol 63:541–556.
599 12. Ryu C-M, Farag MA, Hu C-H, Reddy MS, Wei H-X, Paré PW, Kloepper JW. 2003. Bacterial volatiles promote
600 growth in *Arabidopsis*. Proc Natl Acad Sci U S A 100:4927–4932.
601 13. Haas D, Défago G. 2005. Biological control of soil-borne pathogens by fluorescent pseudomonads. Nat Rev
602 Microbiol 3:307–319.
603 14. Pieterse CMJ, Zamioudis C, Berendsen RL, Weller DM, Van Wees SCM, Bakker PAHM. 2014. Induced
604 systemic resistance by beneficial microbes. Annu Rev Phytopathol 52:347–375.
605 15. Raaijmakers JM, Bonsall RF, Weller DM. 1999. Effect of population density of *Pseudomonas fluorescens* on
606 production of 2,4-Diacetylphloroglucinol in the rhizosphere of wheat. Phytopathology 89:470–475.
607 16. Raaijmakers JM, Leeman M, Oorschot van MM., et.al. 1995. Dose-response relationships in biological
608 control of Fusarium wilt of radish by *Pseudomonas* spp.. Phytopathology 85:1075–1081.
609 17. Lugtenberg BJJ, Dekkers L, Bloemberg G V. 2001. Molecular determinants of rhizosphere colonization
610 *Pseudomonas*. Annu Rev Phytopathol 39:461–490.
611 18. de Weger LA, van der Vlugt CI, Wijfjes AH, Bakker PA, Schippers B, Lugtenberg B. 1987. Flagella of a plant-
612 growth-stimulating *Pseudomonas fluorescens* strain are required for colonization of potato roots. J
613 Bacteriol 169:2769–2773.
614 19. de Weger LA, van Loosdrecht MC, Klaassen HE, Lugtenberg B. 1989. Mutational changes in physiochemical
615 cell surface properties of plant-growth-stimulating *Pseudomonas* spp. do not influence the attachment
616 properties of the cells. J Bacteriol 171:2756–2761.
617 20. Simons M, Permentier HP, de Weger LA, Wijffelman CA, Lugtenberg BJJ. 1997. Amino acid synthesis is
618 necessary for tomato root colonization by *Pseudomonas fluorescens* strain WCS365. Mol Plant-Microbe
619 Interact 10:102–106.
620 21. Berendsen RL, Vismans G, Yu K, Song Y, de Jonge R, Burgman WP, Burmølle M, Herschend J, Bakker PAHM,
621 Pieterse CMJ. 2018. Disease-induced assemblage of a plant-beneficial bacterial consortium. ISME J
622 12:1496–1507.
623 22. Yu K, Liu Y, Tichelaar R, Savant N, Lagendijk E, van Kuijk SJL, Stringlis IA, van Dijken AJH, Pieterse CMJ,
624 Bakker PAHM, Haney CH, Berendsen RL. 2019. Rhizosphere-associated *Pseudomonas* suppress local root
625 immune responses by gluconic acid-mediated lowering of environmental pH. Curr Biol 29:3913-3920.e4.
626 23. Cole BJ, Feltcher ME, Waters RJ, Wetmore KM, Mucyn TS, Ryan EM, Wang G, Ul-Hasan S, McDonald M,
627 Yoshikuni Y, Malmstrom RR, Deutschbauer AM, Dangl JL, Visel A. 2017. Genome-wide identification of

628 bacterial plant colonization genes. PLoS Biol 15:e2002860.
629 24. Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC. 2012. Experimental evolution. Trends Ecol
630 Evol 27:547–560.
631 25. Jousset A, Schuldes J, Keel C, Maurhofer M, Daniel R, Scheu S, Thuermer A. 2014. Full-genome sequence of
632 the plant growth-promoting bacterium *Pseudomonas protegens* CHA0. Genome Announc 2:e00322-14.
633 26. Li E, de Jonge R, Chen L, Friman V, Bakker P, Pieterse C, Jousset A. 2020. Rapid evolution of plant-bacterium
634 mutualism in the rhizosphere. SSRN Electron J https://doi.org/10.2139/ssrn.3650572.
635 27. Oren Y, Smith MB, Johns NI, Kaplan Zeevi M, Biran D, Ron EZ, Corander J, Wang HH, Alm EJ, Pupko T. 2014.
636 Transfer of noncoding DNA drives regulatory rewiring in bacteria. Proc Natl Acad Sci U S A 111:16112–
637 16117.
638 28. Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological
639 evolution. Cell 134:25–36.
640 29. Kupferschmied P, Chai T, Flury P, Blom J, Smits THM, Maurhofer M, Keel C. 2016. Specific surface glycan
641 decorations enable antimicrobial peptide resistance in plant-beneficial *pseudomonads* with insect-
642 pathogenic properties. Environ Microbiol 18:4265–4281.
643 30. Galperin MY, Makarova KS, Wolf YI, Koonin E V. 2015. Expanded Microbial genome coverage and improved
644 protein family annotation in the COG database. Nucleic Acids Res 43:D261–D269.
645 31. Barrett RDH, Maclean RC, Bell G. 2006. Mutations of intermediate effect are responsible for adaptation in
646 evolving *Pseudomonas fluorescens* populations. Biol Lett 2:236–238.
647 32. Allison GE, Verma NK. 2000. Serotype-converting bacteriophages and O-antigen modification in *Shigella*
648 *flexneri*. Trends Microbiol 8:17–23.
649 33. Chang W-S, Park K-M, Koh S-C, So J-S. 2008. Characterization of the *Bradyrhizobium japonicum galE* gene:
650 its impact on lipopolysaccharide profile and nodulation of soybean. FEMS Microbiol Lett 280:242–249.
651 34. Liu B, Knirel YA, Feng L, Perepelov A V., Senchenkova SN, Wang Q, Reeves PR, Wang L. 2008. Structure and
652 genetics of *Shigella* O antigens. FEMS Microbiol Rev 32:627–653.
653 35. Samuel G, Reeves P. 2003. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar
654 precursor synthesis and O-antigen assembly. Carbohydr Res 338:2503–2519.
655 36. Brencic A, McFarland KA, McManus HR, Castang S, Mogno I, Dove SL, Lory S. 2009. The GacS/GacA signal
656 transduction system of *Pseudomonas aeruginosa* acts exclusively through its control over the transcription
657 of the RsmY and RsmZ regulatory small RNAs. Mol Microbiol 73:434–445.
658 37. Brown D. 2010. A mathematical model of the Gac/Rsm quorum sensing network in *Pseudomonas*
659 *fluorescens*. BioSystems 101:200–212.
660 38. Heeb S, Blumer C, Haas D. 2002. Regulatory RNA as mediator in GacA/RsmA-dependent global control of
661 exoproduct formation in *Pseudomonas fluorescens* CHA0. J Bacteriol 184:1046–1056.
662 39. Valverde C, Heeb S, Keel C, Haas D. 2003. RsmY, a small regulatory RNA , is required in concert with RsmZ
663 for GacA-dependent expression of biocontrol traits in *Pseudomonas fluorescens* CHA0. Mol Microbiol
664 50:1361–1379.
665 40. Heeb S, Valverde C, Haas D. 2005. Role of the stress sigma factor RpoS in GacA / RsmA-controlled
666 secondary metabolism and resistance to oxidative stress in *Pseudomonas fluorescens* CHA0. FEMS
667 Microbiol Lett 243:251–258.
668 41. Hassan KA, Johnson A, Shaffer BT, Ren Q, Kidarsa TA, Elbourne LDH, Hartney S, Duboy R, Goebel NC,
669 Zabriskie TM, Paulsen IT, Loper JE. 2010. Inactivation of the GacA response regulator in *Pseudomonas*
670 *fluorescens* Pf-5 has far-reaching transcriptomic consequences. Environ Microbiol 12:899–915.
671 42. Song C, Kidarsa TA, Mortel JE Van De, Loper JE, Jos M. 2016. Living on the edge: emergence of spontaneous
672 gac mutations in *Pseudomonas protegens* during swarming motility. Environ Microbiol 18:3453–3465.
673 43. Hunter P. 2008. The great leap forward. EMBO Rep 9:608–611.
674 44. Zambrano MM, Siegele DA, Almirón M, Tormo A, Kolter R. 1993. Microbial competition: *Escherichia coli*
675 mutants that take over stationary phase cultures. Science 259:1757–1760.
676 45. Sumby P, Whitney AR, Graviss EA, DeLeo FR, Musser JM. 2006. Genome-wide analysis of group A
677 streptococci reveals a mutation that modulates global phenotype and disease specificity. PLoS Pathog 2:e5.
678 46. Giraud A, Arous S, De Paepe M, Gaboriau-Routhiau V, Bambou JC, Rakotobe S, Lindner AB, Taddei F, Cerf-
679 Bensussan N. 2008. Dissecting the genetic components of adaptation of *Escherichia coli* to the mouse gut.
680 PLoS Genet 4:0052–0061.

681 47. Damkiaer S, Yang L, Molin S, Jelsbak L. 2013. Evolutionary remodeling of global regulatory networks during
682 long-term bacterial adaptation to human hosts. Proc Natl Acad Sci U S A 110:7766–7771.
683 48. Sánchez-Contreras M, Martín M, Villacieros M, O'Gara F, Bonilla I, Rivilla R. 2002. Phenotypic selection and
684 phase variation occur during alfalfa root colonization by *Pseudomonas fluorescens* F113. J Bacteriol
685 184:1587–1596.
686 49. Bowden SD, Hale N, Chung JCS, Hodgkinson JT, Spring DR, Welch M. 2013. Surface swarming motility by
687 *Pectobacterium atrosepticum* is a latent phenotype that requires O antigen and is regulated by quorum
688 sensing. Microbiology 159:2375–2385.
689 50. Berry MC, McGhee GC, Zhao Y, Sundin GW. 2009. Effect of a *waaL* mutation on lipopolysaccharide
690 composition, oxidative stress survival, and virulence in *Erwinia amylovora*. FEMS Microbiol Lett 291:80–87.
691 51. Lapouge K, Schubert M, Allain FH-T, Haas D. 2007. Gac/Rsm signal transduction pathway of γ-
692 proteobacteria: from RNA recognition to regulation of social behaviour. Mol Microbiol 67:241–253.
693 52. Matsuyama BY, Krasteva P V., Baraquet C, Harwood CS, Sondermann H, Navarro MVAS. 2016. Mechanistic
694 insights into c-di-GMP-dependent control of the biofilm regulator FleQ from *Pseudomonas aeruginosa*.
695 Proc Natl Acad Sci U S A 113:E209–E218.
696 53. Patrick M, Korotkov K V., Hol WGJ, Sandkvist M. 2011. Oligomerization of EpsE coordinates residues from
697 multiple subunits to facilitate ATPase activity. J Biol Chem 286:10378–10386.
698 54. Caiazza NC, O'Toole GA. 2004. SadB is required for the transition from reversible to irreversible attachment
699 during biofilm formation by *Pseudomonas aeruginosa* PA14. J Bacteriol 186:4476–4485.
700 55. Muriel C, Blanco-Romero E, Trampari E, Arrebola E, Durán D, Redondo-Nieto M, Malone JG, Martín M,
701 Rivilla R. 2019. The diguanylate cyclase AdrA regulates flagellar biosynthesis in *Pseudomonas fluorescens*
702 F113 through SadB. Sci Rep 9:1–9.
703 56. Martínez-Granero F, Navazo A, Barahona E, Redondo-Nieto M, Rivilla R, Martín M. 2012. The Gac-Rsm and
704 SadB signal transduction pathways converge on AlgU to downregulate motility in *Pseudomonas*
705 *fluorescens*. PLoS One 7:e31765.
706 57. Chaparro JM, Badri D V., Bakker MG, Sugiyama A, Manter DK, Vivanco JM. 2013. Root exudation of
707 phytochemicals in *Arabidopsis* follows specific patterns that are developmentally programmed and
708 correlate with soil microbial functions. PLoS One 8:e55731.
709 58. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018.
710 A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol
711 430:2237–2243.
712 59. Rudrappa T, Biedrzycki ML, Kunjeti SG, Donofrio NM, Czymmek KJ, Paré PW, Bais HP. 2010. The
713 rhizobacterial elicitor acetoin induces systemic resistance in *Arabidopsis thaliana*. Commun Integr Biol
714 3:130–138.
715 60. Haas D, Keel C. 2003. Regulation of antibotic production in root-colonizing *Pseudomonas* spp. and
716 relavance for biological control of plant disease. Annu Rev Phytopathol 41:117–153.
717 61. Raaijmakers JM, De Bruijn I, Nybroe O, Ongena M. 2010. Natural functions of lipopeptides from *Bacillus*
718 and *Pseudomonas* : more than surfactants and antibiotics. FEMS Microbiol Rev 34:1037–1062.
719 62. Martínez-Granero F, Rivilla R, Martín M. 2006. Rhizosphere selection of highly motile phenotypic variants
720 of *Pseudomonas fluorescens* with enhanced competitive colonization ability. Appl Environ Microbiol
721 61:1946–1952.
722 63. Romling U, Galperin MY, Gomelsky M. 2013. Cyclic di-GMP: the first 25 years of a universal bacterial
723 second messenger. Microbiol Mol Biol Rev 77:1–52.
724 64. Caiazza NC, Merritt JH, Brothers KM, O'Toole GA. 2007. Inverse regulation of biofilm formation and
725 swarming motility by *Pseudomonas aeruginosa* PA14. J Bacteriol 189:3603–3612.
726 65. Simm R, Morr M, Kader A, Nimtz M, Römling U. 2004. GGDEF and EAL domains inversely regulate cyclic di-
727 GMP levels and transition from sessibility to motility. Mol Microbiol 53:1123–1134.
728 66. Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, D'Argenio DA, Miller SI, Ramsey BW, Speert
729 DP, Moskowitz SM, Burns JL, Kaul R, Olson M V. 2006. Genetic adaptation by *Pseudomonas aeruginosa* to
730 the airways of cystic fibrosis patients. Proc Natl Acad Sci U S A 103:8487–92.
731 67. Winstanley C, O'Brien S, Brockhurst MA. 2016. *Pseudomonas aeruginosa* evolutionary adaptation and
732 diversification in cystic fibrosis chronic lung infections. Trends Microbiol 24:327–337.
733 68. Lerouge I, Vanderleyden J. 2002. O-antigen structural variation: mechanisms and possible roles in

734 animal/plant–microbe interactions. FEMS Microbiol Rev 26:17–47.
735 69. Whitfield C, Trent MS. 2014. Biosynthesis and export of bacterial lipopolysaccharides. Annu Rev Biochem
736 83:99–128.
737 70. Maldonado RF, Sá-Correia I, Valvano MA. 2016. Lipopolysaccharide modification in Gram-negative bacteria
738 during chronic infection. FEMS Microbiol Rev 40:480–493.
739 71. Murray GL, Attridge SR, Morona R. 2006. Altering the length of the lipopolysaccharide O antigen has an
740 impact on the interaction of *Salmonella enterica* serovar Typhimurium with macrophages and
741 complement. J Bacteriol 188:2735–2739.
742 72. Ranf S. 2016. Immune sensing of lipopolysaccharide in plants and animals: same but different. PLoS Pathog
743 12:e1005596.
744 73. Newman M-A, von Roepenack E, Daniels M, Dow M. 2000. Lipopolysaccharides and plant responses to
745 phytopathogenic bacteria. Mol Plant Pathol 1:25–31.
746 74. Di Lorenzo F, Silipo A, Andersen Gersby LB, Palmigiano A, Lanzetta R, Garozzo D, Boyer C, Pruvost O,
747 Newman MA, Molinaro A. 2017. *Xanthomonas citri* pv. *citri* pathotypes: LPS structure and function as
748 microbe-associated molecular patterns. ChemBioChem 18:772–781.
749 75. Newman MA, Daniels MJ, Dow JM. 1995. Lipopolysaccharide from *Xanthomonas campestris* induces
750 defense-related gene expression in *Brassica campestris*. Mol Plant-Microbe Interact 8:778–80.
751 76. Barton-Willis PA, Wang MC, Holliday MJ, Long MR, Keen NT. 1984. Purification and composition of
752 lipopolysaccharides from *Pseudomonas syringae*. pv. *glycinea*. Physiol Plant Pathol 25:387–398.
753 77. Erbs G, Newman M-A. 2003. The role of lipopolysaccharides in induction of plant defence responses. Mol
754 Plant Pathol 4:421–425.
755 78. Leeman M, Pelt JA van, Ouden FM den, Heinsbroek M, Bakker PAHM, Schippers B. 1995. Induction of
756 systemic resistance against *Fusarium wilt* of radish by lipopolysaccharides of *Pseudomonas fluorescens*.
757 Phytopathol 85:1021–1027.
758 79. Van Wees SCM, Pieterse CMJ, Trijssenaar A, Van ’t Westende YAM, Hartog F, Van Loon LC. 1997.
759 Differential induction of systemic resistance in *Arabidopsis* by biocontrol bacteria. Mol Plant-Microbe
760 Interact 10:716–724.
761 80. Nakao R, Senpuku H, Watanabe H. 2006. *Porphyromonas gingivalis galE* is involved in lipopolysaccharide
762 O-antigen synthesis and biofilm formation. Infect Immun 74:6145–53.
763 81. Ramette A, Frapolli M, Saux MF-L, Gruffaz C, Meyer J-M, Défago G, Sutra L, Moënne-Loccoz Y. 2011.
764 *Pseudomonas protegens* sp. nov., widespread plant-protecting bacteria producing the biocontrol
765 compounds 2,4-diacetylphloroglucinol and pyoluteorin. Syst Appl Microbiol 34:180–188.
766 82. Stutz EW, Défago G, Kern H. 1986. Naturally occurring fluorescent pseudomonads involved in suppression
767 of black root rot of tobacco. Phytopathology 76:181–185.
768 83. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for
769 annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of
770 *Drosophila melanogaster* strain *w1118; iso-2; iso-3*. Fly (Austin) 6:80–92.
771 84. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
772 Bioinformatics 26:841–842.
773 85. Højberg O, Schnider U, Winteler H V, Sørensen J, Haas D. 1999. Oxygen-sensing reporter strain of
774 *Pseudomonas fluorescens* for monitoring the distribution of low-oxygen habitats in soil. Appl Environ
775 Microbiol 65:4085–4093.
776 86. Glickmann E, Dessaux Y. 1995. A critical examination of the specificity of the salkowski reagent for indolic
777 compounds produced by phytopathogenic bacteria. Appl Environ Microbiol 61:793–796.
778 87. Alexander DB, Zuberer DA. 1991. Use of chrome azurol S reagents to evaluate siderophore production by
779 rhizosphere bacteria. Biol Fertil Soils 12:39–45.
780 88. Oberhänsli T, Défago G, Haas D. 1991. Indole-3-acetic acid (IAA) synthesis in the biocontrol strain CHA0 of
781 *Pseudomonas fluorescens*: role of tryptophan side chain oxidase. Microbiology 137:2273–2279.
782 89. Moskowitz SM, Foster JM, Emerson J, Burns JL. 2004. Clinically feasible biofilm susceptibility assay for
783 isolates of *Pseudomonas aeruginosa* from patients with cystic fibrosis. J Clin Microbiol 42:1915–1922.
784 90. Christensen PJ, Cook FD. 1972. The isolation and enumeration of cytophagas. Can J Microbiol 18:1933–
785 1940.
786 91. Murtagh F, Legendre P. 2014. Ward’s hierarchical agglomerative clustering method: which algorithms

787 implement ward's criterion? J Classif 31:274–295.
788 92. Scrucca L, Fop M, Murphy TB, Raftery AE. 2016. Mclust 5: clustering, classification and density estimation
789 using gaussian finite mixture models. R J 8:289–317.
790 93. Kai Lee H, Kiat Lee C, Ping Loh T, Wei-Tze Tang J, Anantharajah Tambyah P, Siew-Chuan Koay E. 2011. High-
791 resolution melting approach to efficient identification and quantification of H275Y mutant influenza
792 H1N1/2009 virus in mixed-virus-population samples. J Clin Microbiol 49:3555–3559.
793 94. Zhong W-L, Wang L, Wu X, Zhang J, Chen X-F, Zhang W, Dou X, Yu B. 2016. Development of unlabeled
794 probe based high-resolution melting analysis for detection of filaggrin gene mutation c.3321delA. J Clin Lab
795 Anal 30:892–896.
796 95. Capper RL, Jin YK, Lundgren PB, Peplow LM, Matz M V., van Oppen MJH. 2015. Quantitative high resolution
797 melting: two methods to determine SNP allele frequencies from pooled samples. BMC Genet 16:62.
798