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.

36 ## Introduction

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

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

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

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

85

86

87 ## Results

88 ### Mutational events in independent evolutionary lines

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

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

113 **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

114 [1]Clusters of Orthologous Groups (COGs); https://www.ncbi.nlm.nih.gov/COG/(30)

115 **Identification of potential root colonization genes**

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

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

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

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

158

159 **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	

160 [1]DNA sequence change positions are relative to the cDNA (c); del represents a deletion

161 [2]X represents a stop codon (at its relative position in case of a shifted frame); fs represents a frame shift; del represents a
162 deletion

163 [3]Mutations that are selected for the analysis of their frequency trajectories using RQ-HMR analysis

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

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

180

181 **Adaption driven by global regulators**

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

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

201

202 **Bacterial motility**

203 Bacterial motility is an important trait for rhizosphere competence, mediating colonization of distal parts
204 of the root system (48), and both LPS and the GacS/GacA two-component regulator system affect this
205 trait. The O-antigenic side chain of the LPS was reported to contribute to swimming and swarming
206 motility in the plant pathogenic bacterium *Erwinia amylovora* (49, 50). The GacS/GacA two-component
207 regulator system controls bacterial motility, for example by affecting transcription of genes related to
208 flagella and biosurfactant biosynthesis (42, 48, 51). We also identified several other mutations across
209 the various evolutionary lines that can be linked to bacterial motility. For instance, a disruptive mutation
210 in the flagellar biosynthesis protein-coding gene *FlhA* (H393Q.fsX15) that is involved in the biogenesis of
211 the bacterial flagellum (Table 1) appeared in evolutionary line 4 reaching up to a predicted frequency of
212 75% (Table 2). In *P. fluorescens* Pf-5, and in *P. aeruginosa* PA01 FlhA is reported to be essential for
213 swimming motility (41).

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

226 completely (53). SadB, finally, encodes a HDOD domain-containing protein (Table 1) which shares 77%
227 sequence identity and 99% sequence coverage with *P. aeruginosa* SadB (PA5346). In *P. aeruginosa*, SadB
228 stimulates Pel exopolysaccharide production and the chemotaxis-like cluster CheIV, which in turn affect
229 flagellar motility as well as biofilm formation (54). In *Pseudomonas fluorescens* F113, SadB, together
230 with FleQ control flagellar motility, dependent and independent of the GacS/GacA two-component
231 regulator system (55, 56).

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

243

244 **Dynamics of global phenotypic change**

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

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

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

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

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

283 In contrast to the previous two clusters, traits in clusters 3 (Figure 5d) and 4 (Figure 5e) were more
284 stable, in that most genotypes behave like the ancestor. Only two *gacA* mutants, *gacA*^{E38X} and *gacA*^{Y183S},
285 were associated with a significant decline of bacterial traits associated with biocontrol, such as
286 antifungal activity and biofilm formation, while the *rpoS*^{Q65X} mutation dramatically increased these traits.

287 The double mutation *RS11785*^{S256C} · *flhA*$^{H393Q.fsX15}$, in the background of the disruptive *gacA*^{E38X} mutation
288 (Figure 5e), was the only mutational event that led to a significant increase of general resistance to
289 various environmental stresses. *FlhA*, as previously mentioned encodes the flagellar biosynthesis protein
290 FlhA and is linked to bacterial motility. The LysR-type transcriptional regulator RS11785 in addition to its
291 N-terminal HTH DNA-binding domain (PF00126) contains a predicted C-terminal substrate binding
292 domain (PF03466), which is linked to aromatic compound degradation and resembles that of the type 2
293 periplasmic binding proteins or PBP2s. PBP2s are responsible for binding and uptake of several
294 substrates including polysaccharides and several amino acids. We used HHpred (58) for protein remote
295 homology detection and three-dimensional structure analysis to assess the possible consequence of the
296 S256C amino acid substitution. This analysis identified the remote homolog OxyR from *Escherichia coli*
297 as the best template for modeling (E-value: 2.9e-30, Score: 166.45). Intriguingly, it was previously shown
298 that the activity of OxyR, which represents the master peroxide sensor in gram-negative bacteria, is
299 regulated by intramolecular disulfide-bond formation in the regulatory PBP2-like domain. Notably,
300 RS11785 is unlikely to represent the master peroxide sensor in *P. protegens* CHA0, because another
301 CHA0 protein encoded by *PFLCHA0_RS030065* is much more similar (89% percent pairwise sequence
302 identity versus 29%). Nevertheless, it is tempting to speculate that the here observed S256C substitution
303 might influence regulatory activity by altered intramolecular disulfide-bond formation. Lastly, RS11785
304 bears resemblance (37% sequence identity, 93% sequence coverage) to the LysR-type PBP2-domain
305 containing regulator *alsR* from *Bacillus amyloliquefaciens*. *AlsR* regulates the activity of the acetoin
306 operon (*alsSD*) in response to several signals such as glucose and acetate. Acetoin (3-hydroxy-2-
307 butanone, a precursor for 2,3-butanediol) in turn is known as a potent elicitor of induced systemic
308 resistance (59), and it is linked to general plant growth promotion in *Bacilli*. Given these observations
309 from literature, but in the absence of singular mutations in *flhA* and *RS11785*, the molecular mechanism
310 underlying the observed enhanced resistance in this double mutant remains to be clarified but we
311 hypothesize that altered regulatory activity by RS11785 is a likely driver of the increased stress
312 resistance observed in this mutant.

313 ## Discussion

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

322

323 ### Global regulators and rhizosphere adaptation

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

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

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

351

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

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

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

371 In conclusion, the observed bacterial genetic and phenotypic adaption dynamics emphasize an
372 important role of global regulators, motility, and cell surface structures in bacterial adaptation to its

373 host. The parallel emergence of mutations in similar genes resulting in specific fitness advantages in the
374 rhizosphere, suggest that this evolutionary process is driven by the rhizosphere environment.

375

376 ## Materials and Methods

377

378 ### Experimental setup

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

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

399

400 ### Bacterial life-history traits

401 A variety of bacterial life-history traits, reflecting various aspects of bacterial physiological processes,
402 were measured. We measured bacterial growth yield and resistance to various stresses, including acidic

403 (pH = 5) and alkaline (pH = 9) conditions, oxidative stress in 0.0025% H_2O_2, water potential stress (15%
404 polyethylene glycol (PEG)-6000)), and salt stress (2% NaCl), and resistance to the antibiotics
405 streptomycin (1 µg · ml^{-1}), tetracycline (1 µg · ml^{-1}), and penicillin (5 µg · ml^{-1}). Bacterial growth yield was
406 quantified in modified Ornston and Stanier (OS) minimal medium (85) supplemented with single carbon
407 sources that have been reported to be abundant in Arabidopsis root exudates (Chaparro et al 2013).
408 These included the following carbon sources; alanine, arabinose, butyrolactam, fructose, galactose,
409 glucose, glycerol, glycine, lactic acid, putrescine, serine, succinic acid, threonine, and valine that were
410 added to a final concentration of 0.5 g · l^{-1}. Growth yield was measured as OD_{600} (SPECTROstar Nano
411 plate reader), and this measure was used to determine carbon source utilization ability. In addition, we
412 measured bacterial auxin (indole-3-acetic acid or IAA) production with a colorimetric test (86), iron-
413 chelating ability using a modified chrome azurol S (CAS) assay (87), proteolytic activity by the adapted
414 assay from Smeltzer et al. (1993), tryptophan side chain oxidase activity using a colorimetric assay (88),
415 and biofilm formation using a modified Chrystal Violet staining assay (89). We further assessed bacterial
416 antimicrobial activity by quantifying their effect on growth of the fungi *Verticillium dahliae* and *Fusarium*
417 *oxysporum,* and the bacterium *Ralstonia solanacearum*. Bacterial motility assays were performed on
418 Cook's Cytophaga (CC) medium containing 0.2% tryptone (90). This medium was solidified with 0.3 or
419 0.5% agar to determine swimming and swarming motility, respectively (42). Bacteria were first cultured
420 in liquid CC medium overnight at 28 °C, and subsequently inoculated in the middle of the motility plates
421 using toothpicks. Plates were incubated upside-down at 28 °C for 24 h, and the colony areas were
422 calculated using Image J as a measure of bacterial motility.

423

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

425 *Hierarchical clustering*

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

429 *Model-based clustering*

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

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

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

442

443 **Genotype-phenotype association analysis**

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

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

453

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

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

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

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

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

484

485 **Data availability**

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

490

Acknowledgments

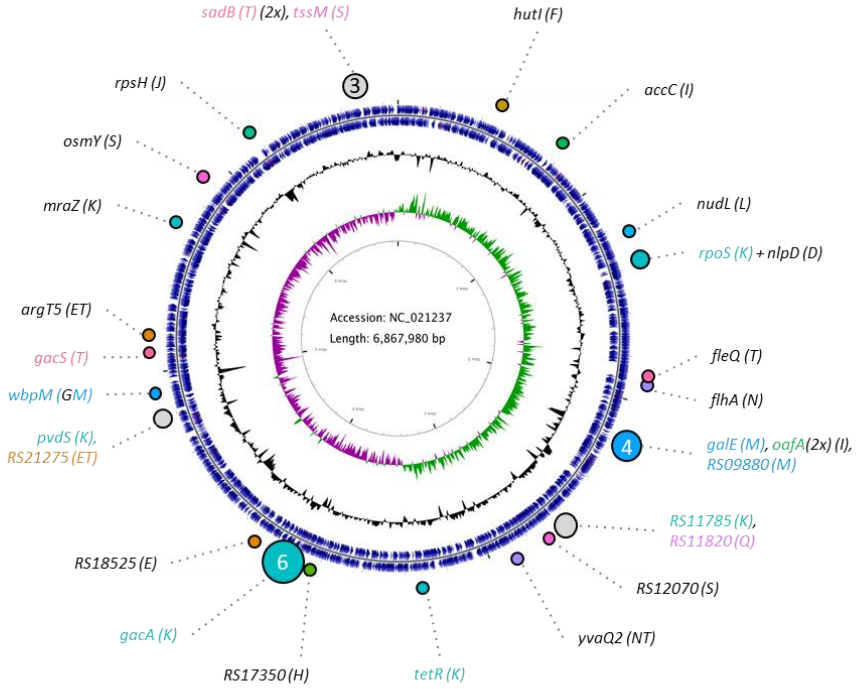
This work was supported by a China Scholarship Council fellowship (to E.L.), a postdoctoral fellowship of the Research Foundation Flanders (FWO 12B8116RN) (to R.D.J.), and a European Research Council Advanced Grant 269072 (to C.M.J.P.).

Author contributions

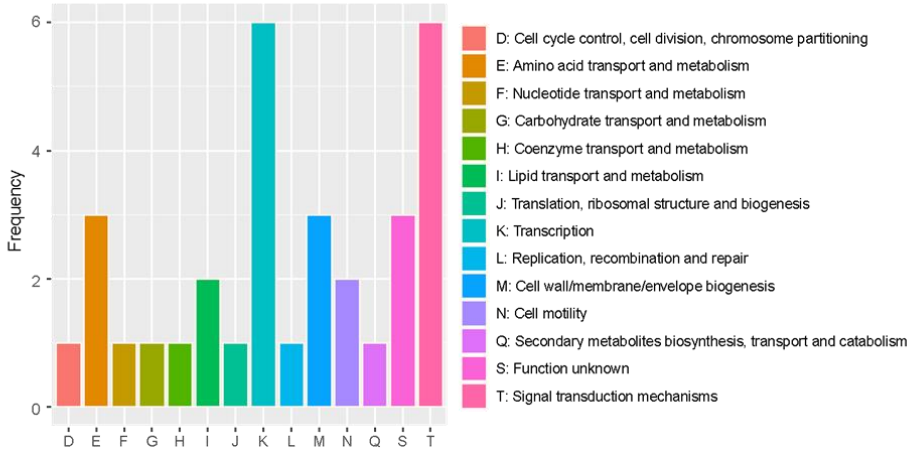
E.L., A.J., C.M.J.P., P.A.H.M.B., and R.D.J. designed experiments; E.L. and H.J. performed experiments; E.L. and R.D.J. analyzed data. All authors discussed the results and contributed to the final manuscript.

500 **Figures and Figure legends**

a)



b)



501

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

508

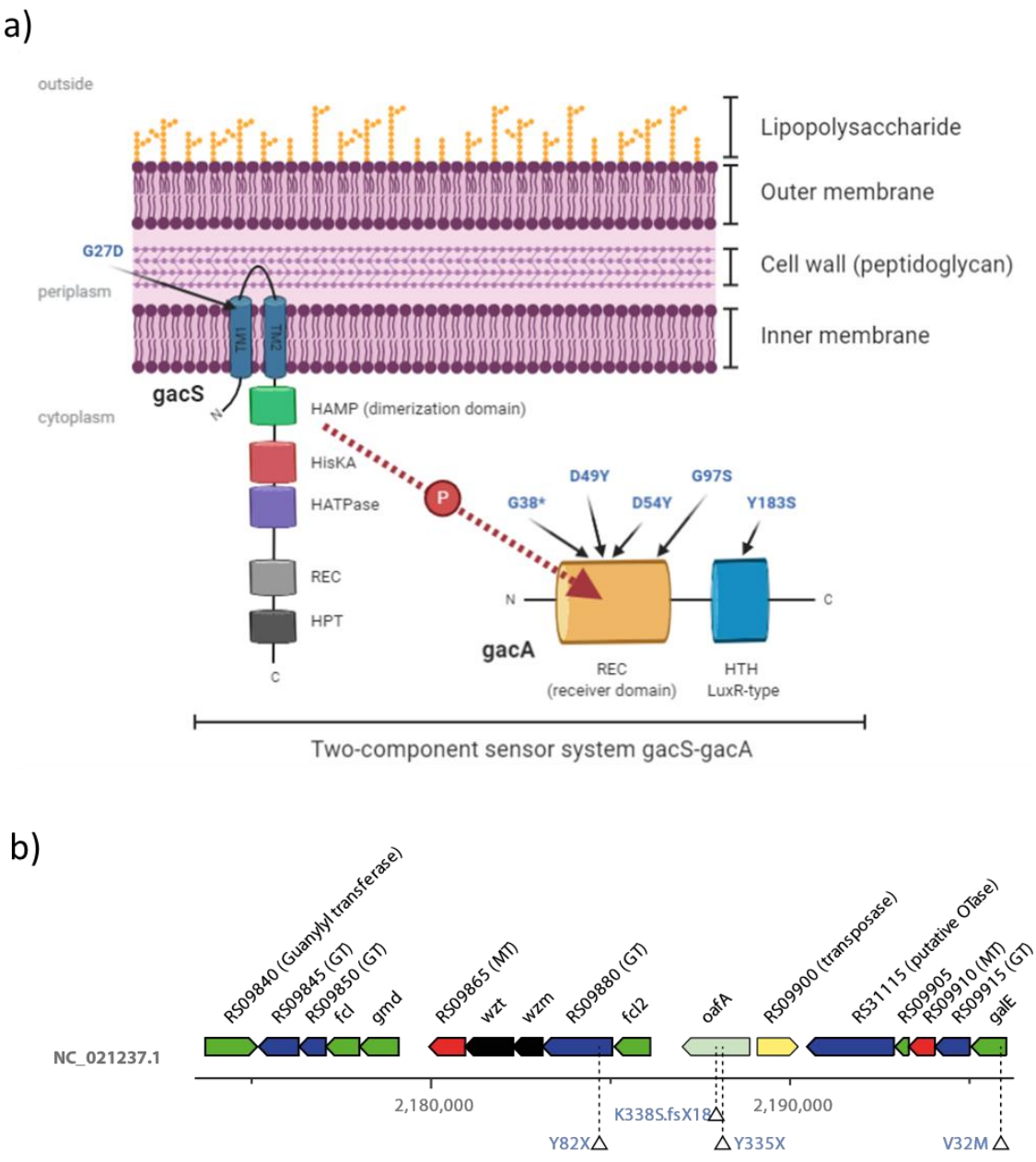


Figure 2. Localization of *gac* and *OBC3* mutations identified in the evolutionary experiment. **a)** Schematic view of the cell wall of the gram-negative bacterium *Pseudomonas protegens* highlighting the presence of the lipopolysaccharide (LPS) on the outside and the presence of the two-component 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).

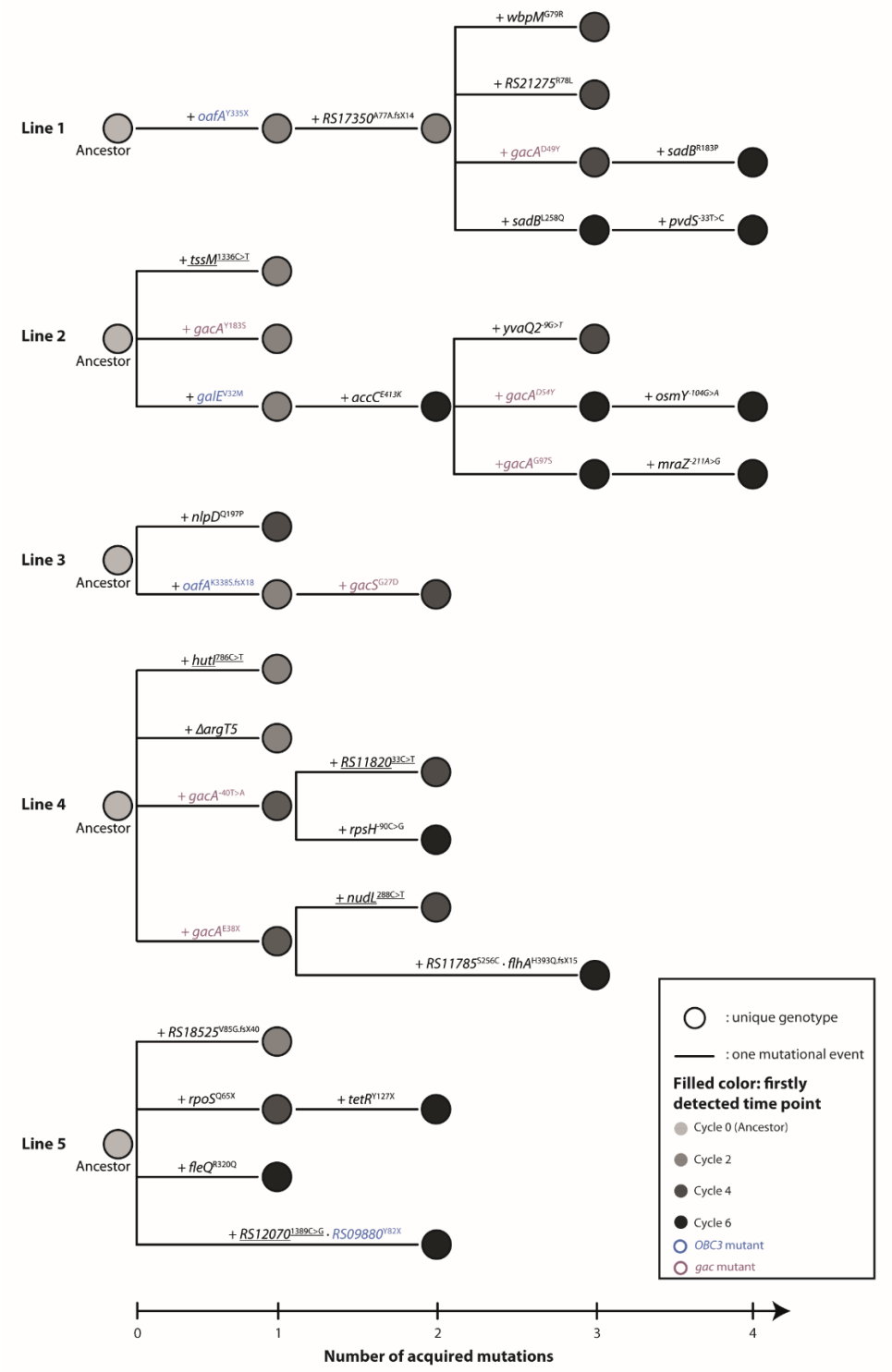


Figure 3. Phylogenetic trees of five independently evolving *P. protegens* CHA0 populations. Phylogenies for 18 genomes from each population (16 for the population in experimental line 4), based on the sequential appearance of mutations. Synonymous mutations are underlined. Circles represent unique bacterial 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.

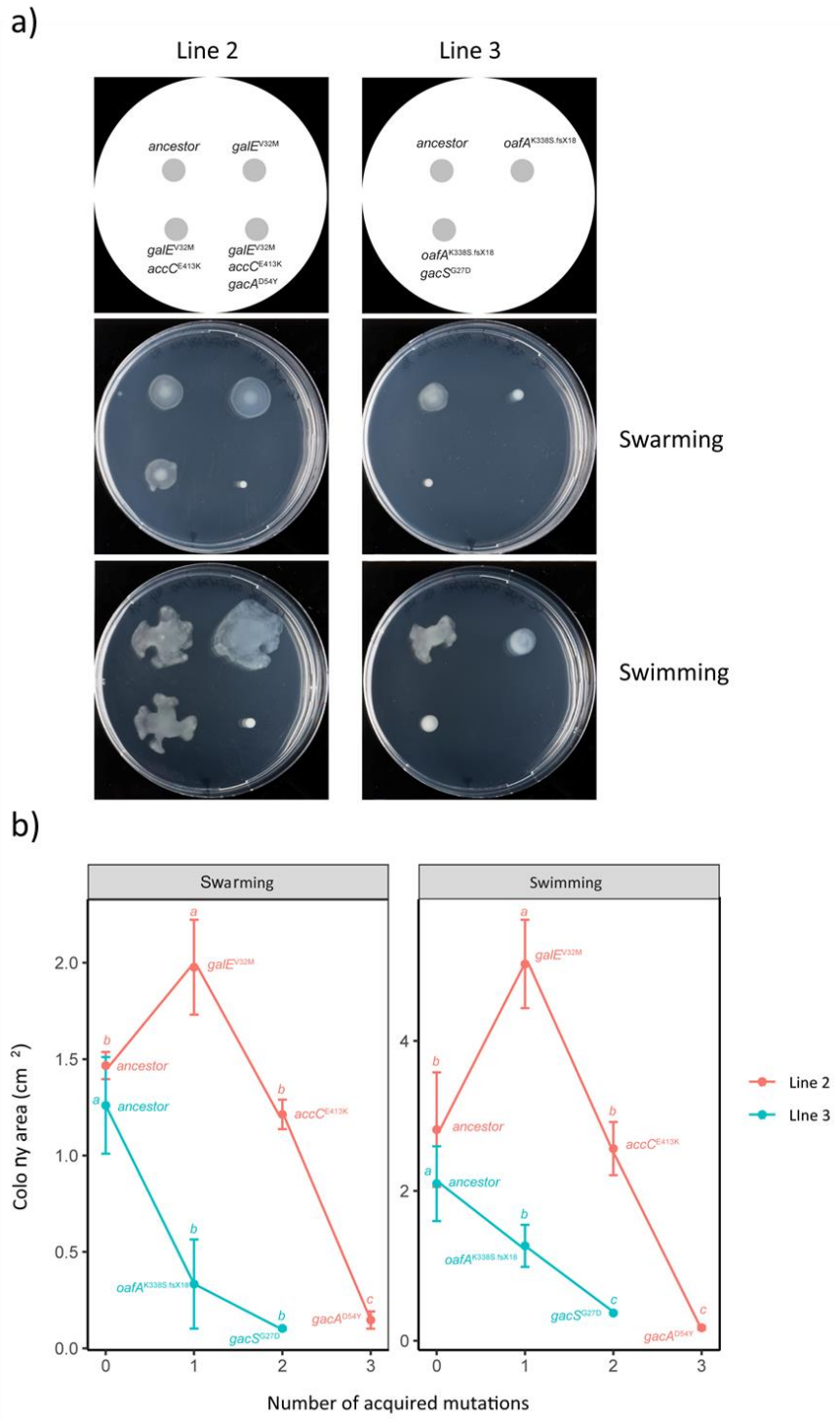
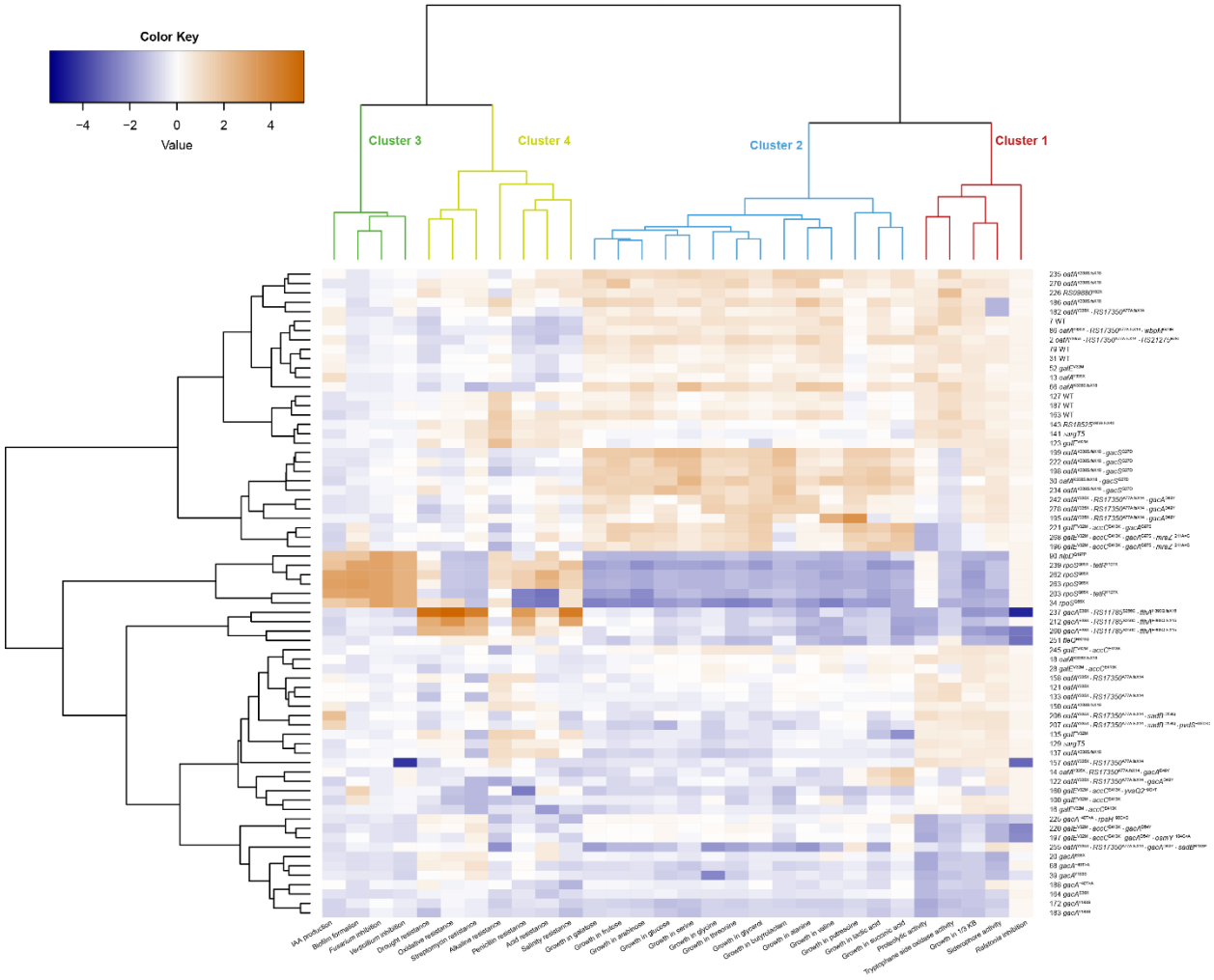


Figure 4. Effects of *gacS/gacA* mutations on bacterial swimming and swarming behaviour. **a)** The top panels illustrate the experimental layout of motility assays, that are pictured in the middle and bottom panels for swarming and swimming respectively. **b)** The accumulative effects of mutations 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
Fusarium inhibition
Verticillium inhibition
Drought resistance
Oxidative resistance
Streptomycin resistance
Alkaline resistance
Penicillin resistance
Acid resistance
Salinity resistance
Growth in galactose
Growth in fructose
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 putrescine
Growth in lactic acid
Growth in succinic acid
Proteolytic activity
Tryptophane side oxidase activity
Growth in 1/3 KB
Siderophore activity
Rotoxivia inhibition

531

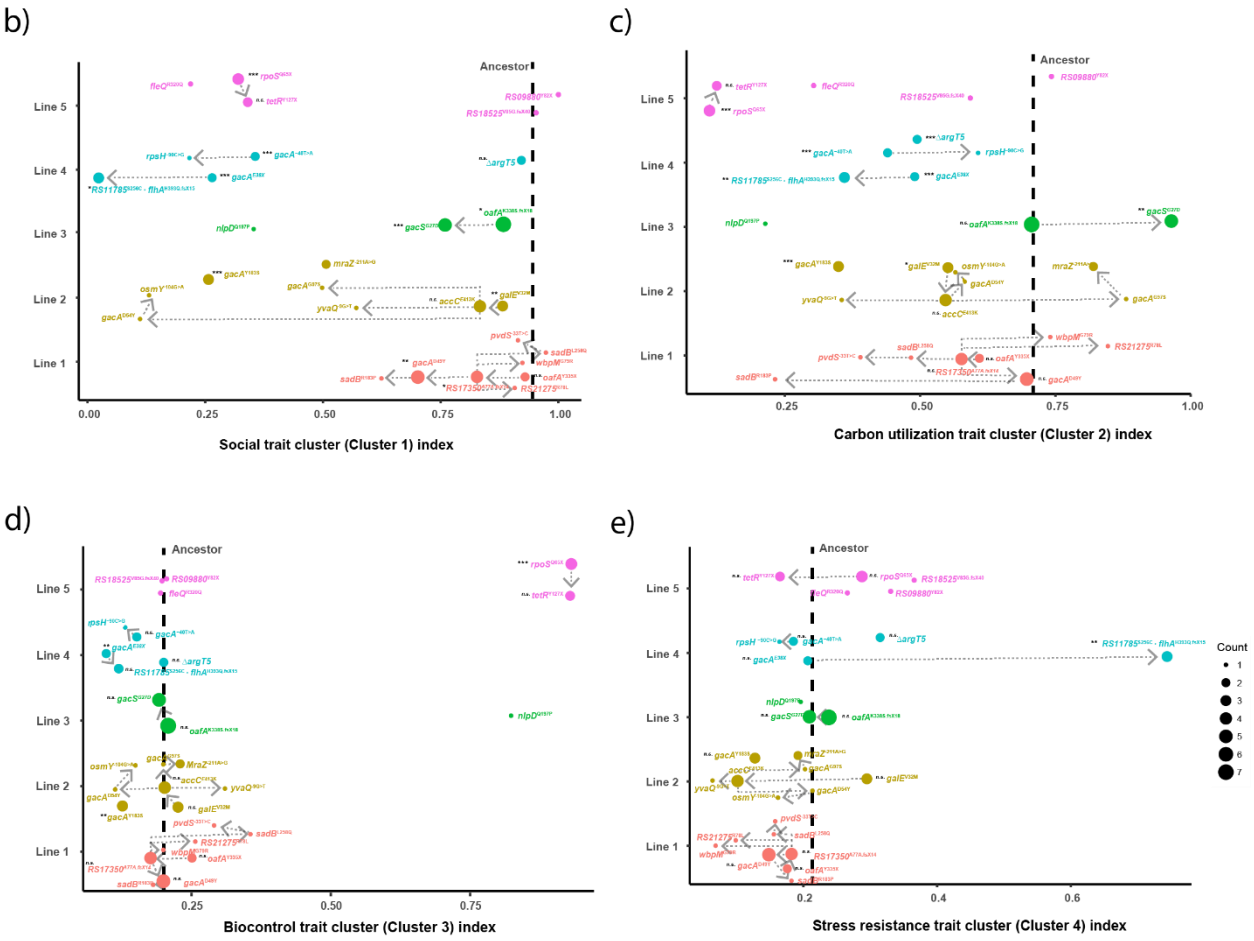
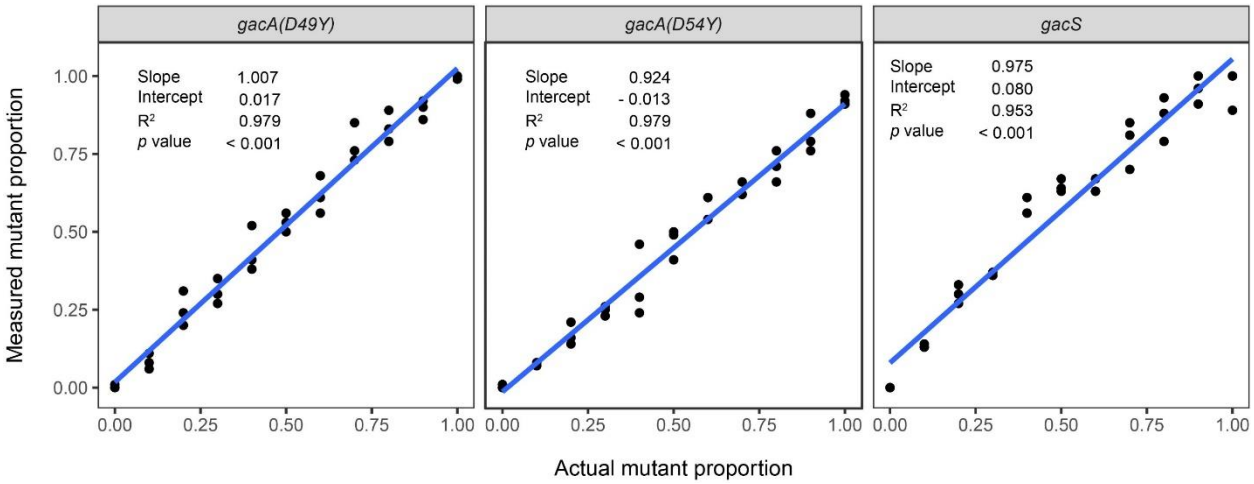


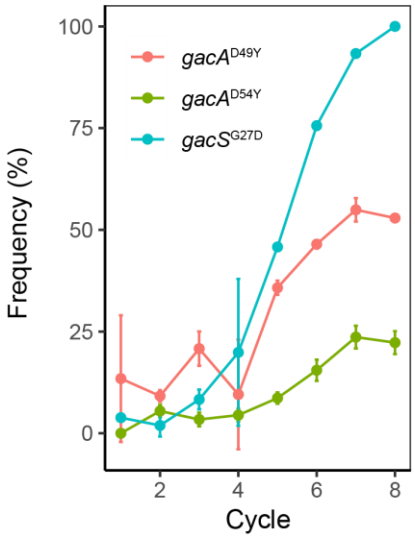
Figure 5. Genotype-phenotype association analysis of evolved bacterial populations. **a)** Heatmap representing the performance of 30 life-history traits of six ancestral and 63 evolved isolates (isolates without mutations or only with synonymous mutation were excluded for this association analysis). 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.

544 ## Supplementary Figures:



545

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



552

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

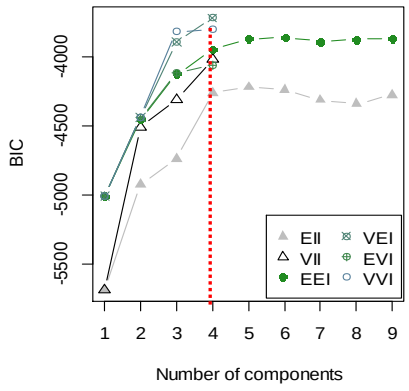
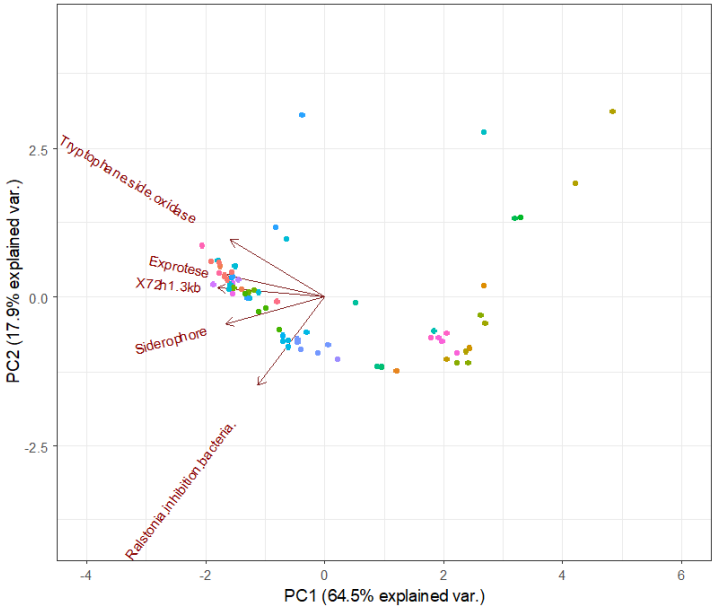
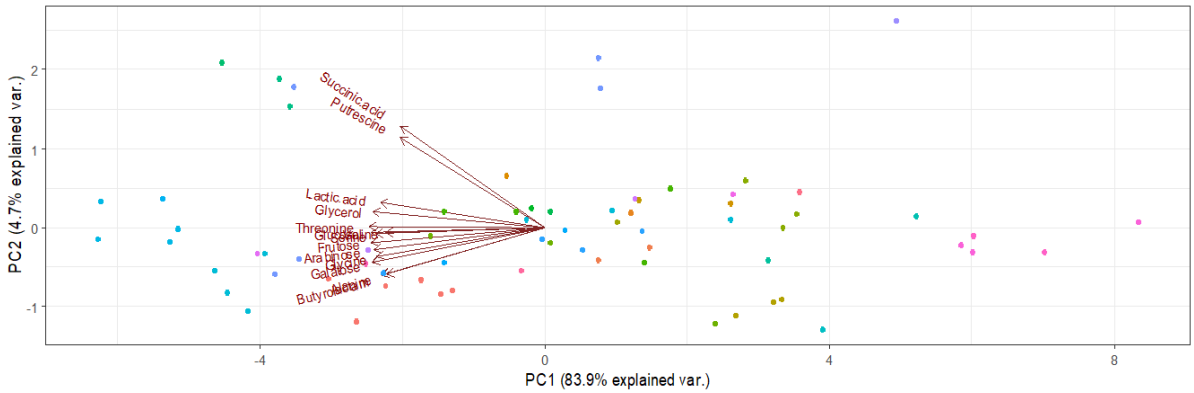


Figure S3. Selection of the best clustering model using Bayesian Information Criterion (BIC) using the *mclust* R package. A large BIC score indicates strong evidence for the corresponding model. The VEI model with 4 components best fits our data. VEI indicates that the clusters have variable volume,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

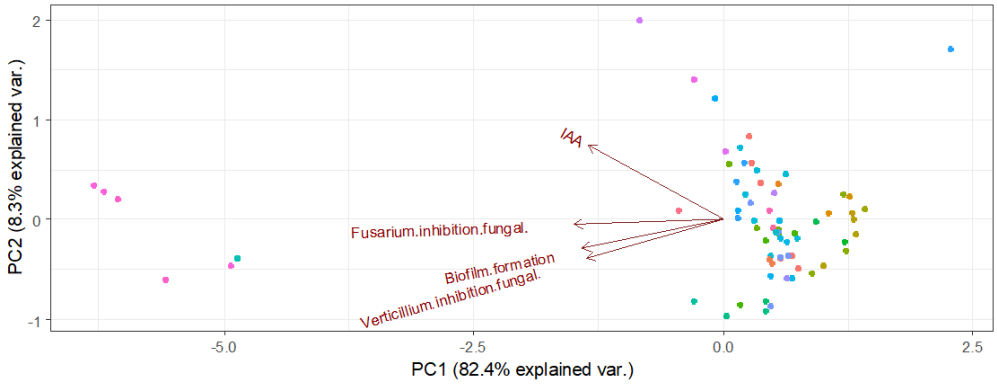
a



b



c



d

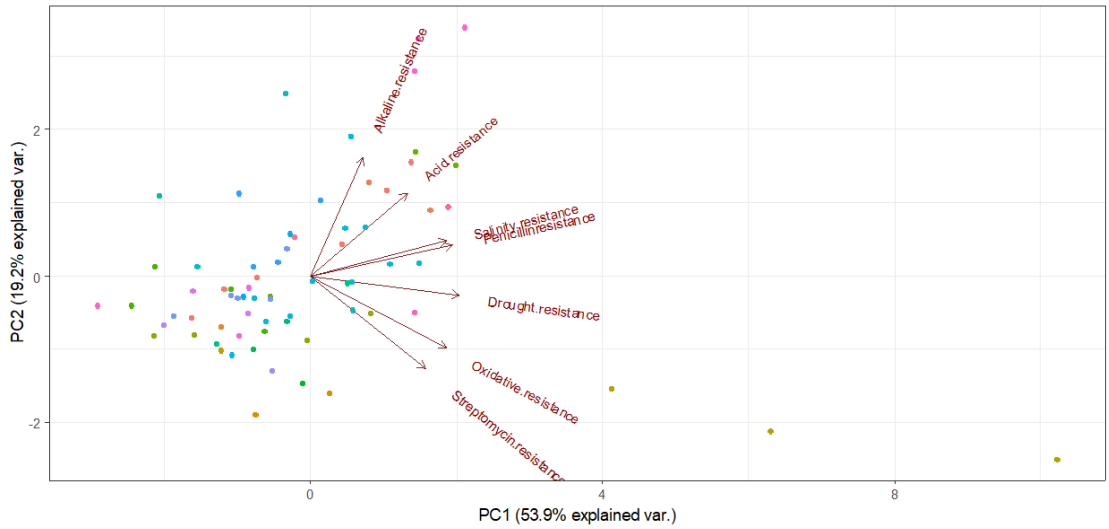


Figure S4. Coordinate axes transformation of each model predicted cluster using principal component analysis (PCA). **a.** PCA plot of social trait cluster (cluster 1); **b.** PCA plot of carbon utilization trait cluster (cluster 2); **c.** PCA

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

571 **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*

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

574

575 **Table S2** Primers and probes used for high-resolution melting (HRM) analysis. For the two *gacA* mutants the same set of
576 primers was used. Underlined bases indicate the position of the single nucleotide point (SNP) mutations within the probe

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

577 sequences. ΔTm (°C) indicates the melting temperature difference between WT-probe duplex and mutant-probe duplex.

578

579 ## References

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

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

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

737 animal/plant–microbe interactions. FEMS Microbiol Rev 26:17–47.
738 69. Whitfield C, Trent MS. 2014. Biosynthesis and export of bacterial lipopolysaccharides. Annu Rev Biochem
739 83:99–128.
740 70. Maldonado RF, Sá-Correia I, Valvano MA. 2016. Lipopolysaccharide modification in Gram-negative bacteria
741 during chronic infection. FEMS Microbiol Rev 40:480–493.
742 71. Murray GL, Attridge SR, Morona R. 2006. Altering the length of the lipopolysaccharide O antigen has an
743 impact on the interaction of *Salmonella enterica* serovar Typhimurium with macrophages and
744 complement. J Bacteriol 188:2735–2739.
745 72. Ranf S. 2016. Immune sensing of lipopolysaccharide in plants and animals: same but different. PLoS Pathog
746 12:e1005596.
747 73. Newman M-A, von Roepenack E, Daniels M, Dow M. 2000. Lipopolysaccharides and plant responses to
748 phytopathogenic bacteria. Mol Plant Pathol 1:25–31.
749 74. Di Lorenzo F, Silipo A, Andersen Gersby LB, Palmigiano A, Lanzetta R, Garozzo D, Boyer C, Pruvost O,
750 Newman MA, Molinaro A. 2017. *Xanthomonas citri* pv. *citri* pathotypes: LPS structure and function as
751 microbe-associated molecular patterns. ChemBioChem 18:772–781.
752 75. Newman MA, Daniels MJ, Dow JM. 1995. Lipopolysaccharide from *Xanthomonas campestris* induces
753 defense-related gene expression in *Brassica campestris*. Mol Plant-Microbe Interact 8:778–80.
754 76. Barton-Willis PA, Wang MC, Holliday MJ, Long MR, Keen NT. 1984. Purification and composition of
755 lipopolysaccharides from *Pseudomonas syringae*. pv. *glycinea*. Physiol Plant Pathol 25:387–398.
756 77. Erbs G, Newman M-A. 2003. The role of lipopolysaccharides in induction of plant defence responses. Mol
757 Plant Pathol 4:421–425.
758 78. Leeman M, Pelt JA van, Ouden FM den, Heinsbroek M, Bakker PAHM, Schippers B. 1995. Induction of
759 systemic resistance against *Fusarium wilt* of radish by lipopolysaccharides of *Pseudomonas fluorescens*.
760 Phytopathol 85:1021–1027.
761 79. Van Wees SCM, Pieterse CMJ, Trijssenaar A, Van 't Westende YAM, Hartog F, Van Loon LC. 1997.
762 Differential induction of systemic resistance in *Arabidopsis* by biocontrol bacteria. Mol Plant-Microbe
763 Interact 10:716–724.
764 80. Nakao R, Senpuku H, Watanabe H. 2006. *Porphyromonas gingivalis galE* is involved in lipopolysaccharide
765 O-antigen synthesis and biofilm formation. Infect Immun 74:6145–53.
766 81. Ramette A, Frapolli M, Saux MF-L, Gruffaz C, Meyer J-M, Défago G, Sutra L, Moënne-Loccoz Y. 2011.
767 *Pseudomonas protegens* sp. nov., widespread plant-protecting bacteria producing the biocontrol
768 compounds 2,4-diacetylphloroglucinol and pyoluteorin. Syst Appl Microbiol 34:180–188.
769 82. Stutz EW, Défago G, Kern H. 1986. Naturally occurring fluorescent pseudomonads involved in suppression
770 of black root rot of tobacco. Phytopathology 76:181–185.
771 83. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for
772 annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of
773 *Drosophila melanogaster* strain *w1118; iso-2; iso-3*. Fly (Austin) 6:80–92.
774 84. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
775 Bioinformatics 26:841–842.
776 85. Højberg O, Schnider U, Winteler H V, Sørensen J, Haas D. 1999. Oxygen-sensing reporter strain of
777 *Pseudomonas fluorescens* for monitoring the distribution of low-oxygen habitats in soil. Appl Environ
778 Microbiol 65:4085–4093.
779 86. Glickmann E, Dessaux Y. 1995. A critical examination of the specificity of the salkowski reagent for indolic
780 compounds produced by phytopathogenic bacteria. Appl Environ Microbiol 61:793–796.
781 87. Alexander DB, Zuberer DA. 1991. Use of chrome azurol S reagents to evaluate siderophore production by
782 rhizosphere bacteria. Biol Fertil Soils 12:39–45.
783 88. Oberhänsli T, Défago G, Haas D. 1991. Indole-3-acetic acid (IAA) synthesis in the biocontrol strain CHA0 of
784 *Pseudomonas fluorescens*: role of tryptophan side chain oxidase. Microbiology 137:2273–2279.
785 89. Moskowitz SM, Foster JM, Emerson J, Burns JL. 2004. Clinically feasible biofilm susceptibility assay for
786 isolates of *Pseudomonas aeruginosa* from patients with cystic fibrosis. J Clin Microbiol 42:1915–1922.
787 90. Christensen PJ, Cook FD. 1972. The isolation and enumeration of cytophagas. Can J Microbiol 18:1933–
788 1940.
789 91. Murtagh F, Legendre P. 2014. Ward's hierarchical agglomerative clustering method: which algorithms

790 implement ward's criterion? J Classif 31:274–295.
791 92. Scrucca L, Fop M, Murphy TB, Raftery AE. 2016. Mclust 5: clustering, classification and density estimation
792 using gaussian finite mixture models. R J 8:289–317.
793 93. Kai Lee H, Kiat Lee C, Ping Loh T, Wei-Tze Tang J, Anantharajah Tambyah P, Siew-Chuan Koay E. 2011. High-
794 resolution melting approach to efficient identification and quantification of H275Y mutant influenza
795 H1N1/2009 virus in mixed-virus-population samples. J Clin Microbiol 49:3555–3559.
796 94. Zhong W-L, Wang L, Wu X, Zhang J, Chen X-F, Zhang W, Dou X, Yu B. 2016. Development of unlabeled
797 probe based high-resolution melting analysis for detection of filaggrin gene mutation c.3321delA. J Clin Lab
798 Anal 30:892–896.
799 95. Capper RL, Jin YK, Lundgren PB, Peplow LM, Matz M V., van Oppen MJH. 2015. Quantitative high resolution
800 melting: two methods to determine SNP allele frequencies from pooled samples. BMC Genet 16:62.
801