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

Running title: *Pseudomonas protegens* adaptation to the rhizosphere

Erqin Li[1,†, #,*], Hao Zhang[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.

[*]E.L. and H.Z. contributed equally to this article. The author order was determined by contribution to writing the manuscript.

[†]Corresponding authors: Ronnie de Jonge (r.dejonge@uu.nl), Erqin Li (erqinli22@gmail.com).

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

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 competition with neighboring microorganisms and host immunity. Forward and reverse genetics has led to the identification of mechanisms that are used by beneficial microorganisms to overcome these challenges such as the production of iron-chelating compounds, the formation of strong biofilms, or the concealment of characteristic microbial molecular patterns that trigger the host immune system. However, how such mechanisms arose from an evolutionary perspective is much less understood. To study bacterial adaptation in the rhizosphere, we employed experimental evolution to track the physiological and genetic dynamics of root-dwelling *Pseudomonas protegens* in the *Arabidopsis thaliana* rhizosphere under axenic conditions. This simplified binary one plant-one bacterium system allows for the amplification of key adaptive mechanisms for bacterial rhizosphere colonization. We identified 35 mutations, including single-nucleotide polymorphisms, insertions, and deletions, distributed over 28 genes. We found that mutations in genes encoding global regulators, and in genes for siderophore production, cell surface decoration, attachment, and motility accumulated in parallel, underlining that bacterial adaptation to the rhizosphere follows multiple strategies. Notably, we observed that motility increased in parallel across multiple independent evolutionary lines. Altogether these results underscore the strength of experimental evolution to identify key genes, pathways, and processes for bacterial rhizosphere colonization, and a methodology for

the development of elite beneficial microorganisms with enhanced root-colonizing capacities that can support sustainable agriculture in the future.

Importance

Beneficial root-associated microorganisms carry out many functions that are essential for plant performance. Establishment of a bacterium on plant roots, however, requires overcoming many challenges. Previously, diverse mechanisms that are used by beneficial microorganisms to overcome these challenges were identified. However, how such mechanisms have developed from an evolutionary perspective is much less understood. Here, we employed experimental evolution to track the evolutionary dynamics of a root-dwelling pseudomonad on the root of Arabidopsis. We find that mutations in global regulators, as well as in genes for siderophore production, cell surface decoration, attachment, and motility accumulate in parallel, underlining these strategies for bacterial adaptation to the rhizosphere. We identified 35 mutations distributed over 28 genes. Altogether our results demonstrate the power of experimental evolution to identify key pathways for rhizosphere colonization and a methodology for the development of elite beneficial microorganisms that can support sustainable agriculture.

Introduction

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

Among all root-dwelling organisms, fluorescent *Pseudomonas* spp. are well characterized in terms of the traits required for their growth in the rhizosphere and for the establishment of beneficial plant-microbe interactions (17, 18). To study bacterial traits involved in efficient root colonization, mutants defective in specific traits suspected to be involved in colonization, are

78 compared to the parental strains for their ability to colonize plant roots. Such studies have
79 highlighted a range of bacterial traits involved in efficient root colonization, including flagella
80 (19), surface lipopolysaccharides (LPS) (20), and amino acid synthesis (21). Using random
81 mutagenesis and by determining the fitness of each mutant in competition with its parental
82 strain in the rhizosphere, many other important traits for rhizosphere competence in
83 *Pseudomonas* were discovered (17). Recently, random mutagenesis in *Pseudomonas*
84 *capeferrum* WCS358 led to the identification of two genes that are important for gluconic acid
85 (GA) biosynthesis. GA, in turn, is essential for the suppression of local, flagellin-induced root
86 immune responses (22). Such suppression was shown to be important for rhizosphere
87 competence as GA-deficient mutants maintain reduced populations in the rhizosphere (22). In
88 another recent study, genome-wide saturation mutagenesis of *Pseudomonas simiae* WCS417r
89 revealed that 2% of the protein-coding genes are important for successful root colonization
90 (23). Mutations that negatively affect rhizosphere competence, and mutations that confer a
91 competitive advantage were identified in this study. The identification of mutations that can
92 lead to increased root colonization (23), suggests that there is room for improvement of
93 bacterial fitness in the rhizosphere.

94 In the present study we used an experimental evolution approach (24) to study how
95 *Pseudomonas protegens* CHA0 (CHA0) evolves during repeated colonization of the rhizosphere
96 of the model plant *Arabidopsis thaliana* (Arabidopsis). The model biological control agent CHA0
97 displays broad-spectrum antagonistic activity against several plant pathogenic fungi and
98 bacteria (11), and its complete genome sequence is available (25). We performed highly
99 controlled experimental evolution in a gnotobiotic and carbon-free sand system in which

bacteria depend solely on the plant for supply of carbon. Following inoculation and establishment on the roots, bacterial populations were transferred to new plants, and this cycle was repeated eight times. We hypothesized that the repeated colonization and establishment of the bacterial population on the plant root would create an environment in which selection pressure drives the accumulation of better colonizers. *In vitro* characterization of individual bacterial colonies from these populations combined with sequencing analysis led to the identification of several evolutionary trajectories involving 35 distinct mutations that impact social traits representing inter-population communication and cooperation, carbon source utilization, motility, or biocontrol activity. By combining experimental evolution with whole genome re-sequencing, we created a powerful screen for the identification of adaptive mutations with positive effects on rhizosphere colonization.

Results

Mutational events in independent evolutionary lines

We previously studied five experimental evolutionary populations, referred to as lines, of CHA0 evolving in the rhizosphere of Arabidopsis in a gnotobiotic system. Independent populations were introduced on the roots and after four weeks of plant growth the populations were transferred to new plants. This cycle of transferring was repeated eight times, and we performed extensive characterizations up until cycle six to account for feasibility (26). In short, for each line, after every cycle we plated a fraction of the population on culture media and randomly picked 16 colonies for extensive phenotypic assessment of bacterial life history traits (26). In order to study adaptation at the genetic level, we selected six colonies from each line at cycles 2, 4, and 6 such that they represented most of the observed phenotypic diversity among the 16 colonies that were initially picked and characterized. These colonies, as well as six colonies from the ancestral population that was initially introduced on the roots, were re-sequenced to an average depth of 25-fold coverage (minimum 10, maximum 70) and used for the identification of single nucleotide polymorphisms (SNPs), as well as small and large insertions or deletions (INDELs). In total we thus set out to acquire genetic data for 96 bacterial colonies (5 lines * 3 cycles * 6 colonies + 6 ancestral colonies). Unfortunately, we were unable to acquire sufficient sequencing data for two colonies from line 4 at cycle 4, yielding a final set of 94 (88 evolved, 6 ancestral) re-sequenced colonies. The six ancestral colonies were all identical, indicating that there was no genetic variation in the starting population and that all observed mutations are *de novo* mutations. In total, one or more mutations were detected in 64 evolved colonies, which collectively represent 73% of the 88 characterized bacterial colonies

(Table S1). We identified 5 synonymous substitutions, 20 nonsynonymous substitutions, and 4 deletions ranging in length from 1 base pair (bp) to about 400 bp, distributed over 22 genes, and 6 additional mutational events located in intergenic regions (Table 1). Mutations located in intergenic regions possibly affect transcription of nearby genes via affecting their regulatory proteins binding sites and subsequent changes in their promoter activity (27, 28). Several mutations were found to be clustered in select genes and/or regions in the CHA0 genome, *e.g.* those in the response regulator *gacA* gene (PFLCHA0_RS17965; *NC_021237:4,039,113-4,039,754*) or in the *OBC3* gene cluster involved in LPS biosynthesis (*NC_021237:2,173,707-2,196,028*)(29), but the majority of the mutations were spread across the 6.1 Mbp CHA0 genome (Figure 1). Functional characterization of the mutated genes by analyzing their Clusters of Orthologous Groups (COGs) annotation revealed that the majority of these genes are involved in transcription (COG term 'K'), signal transduction mechanisms (COG term 'T'), amino acid transport and metabolism (COG term 'E') and cell wall/membrane/envelope biogenesis (COG term 'M') (Figure 1, Table 1).

Identification of potential root colonization genes

Bacterial genes that are involved in colonization of plant roots can be revealed by identifying beneficial mutations that evolve during adaption of bacteria to the rhizosphere environment and have positive effects on root colonization. Over time, such mutants will outcompete the ancestral strain and become dominant in the bacterial population. The observation that only a limited number of mutations accumulated relative to the total number of genes within the

154 genome across the entire experiment, makes it highly unlikely that the same gene would
155 acquire several changes by chance in independent evolutionary lines. Nevertheless, we observe
156 recurrent mutations in several of the same genes and/or pathways (Table 1), which is a strong
157 indication for adaptive evolution. Genes or pathways that acquired mutations in multiple
158 independent CHA0 populations included *gacA* and *gacS*, the *OBC3* gene cluster (*oafA*; *galE*;
159 *PFLCHA0_RS09880* (29)), and a putative pyoverdine siderophore biosynthesis cluster (*pvdS*;
160 *PFLCHA0_RS21275*). Other genes, like *sadB*, were targeted more than once in the same
161 population. Because these genes were repeatedly identified in the CHA0 evolution experiment,
162 they can be assumed to contribute significantly to bacterial fitness in the rhizosphere (Table 1,
163 Figure 2). Moreover, it suggests that the independent evolutionary lines converge on similar
164 evolutionary trajectories involving overlapping biological processes and molecular mechanisms.

165 To illustrate the diverse evolutionary trajectories, we constructed phylogenetic trees of each
166 replicate population (line) using all mutations that were detected (Figure 3). In each line, both
167 the number and depth of branches are very different (Figure 3, Table 2). In line 1 and 3, the
168 populations were swept early by *oafA* mutants and later by specific *gac* mutants, while in line 2
169 and 4, co-existence of multiple genotypes characterized by *gac* mutations was observed
170 indicative of clonal interference between lineages (Figure 3; (31)). Notably, the relative
171 abundance of each genotype in the population is estimated from a small set of sequenced
172 individuals and therefore is expected to lack precision. The evolved isolates contain one to four
173 mutations in general, and in each line, 3 (line 3) to 9 (line 2 and 4) mutations were identified in
174 total. All mutations were unique for each line, as is expected for independently evolving
175 populations.

As the frequency of each mutation (Figure 3, Table 2) was determined from only six bacterial colonies that were isolated and sequenced from each evolutionary line at cycles 2, 4 and 6, we further investigated and accurately measured the population-level frequency at the end of each experimental cycle. We determined the frequency of three *gac* mutations, *i.e.*, *gacA*^{D49Y}, *gacA*^{D54Y}, and *gacS*^{G27D}, with increased accuracy, *i.e.* by PCR-based high-resolution melting (HRM) analysis incorporating mutation-specific HRM probes (Table S2), and increased sampling depth, *i.e.*, at the end of experimental cycles 1 to 8. The HRM methodology allows for the accurate quantification of mutant frequency across a wide range for all three mutations (p<0.001; Figure S1). Using this method, we found that the population-level frequency of these three mutations in the respective evolutionary lines corresponded remarkably well with the frequency previously obtained from the cultured and sequenced isolates (Table 2, Figure S2). These findings corroborate the culture-based quantification of mutant frequency suggesting they provide a reasonable measure for population-level frequency. Intriguingly, it can be seen that the *gacS*^{G27D} mutation reaches fixation after eight experimental cycles, while the *gacA*^{D49Y} and *gacA*^{D54Y} mutants stabilize at around 50% and 25%, respectively (Figure S2). Stabilization of mutations is indicative of frequency-dependent (FD) selection putatively reinforced by clonal interference with co-existing lineages carrying beneficial mutations in *sadB* (L258Q) and *gacA* (G97S) (Figure 3). FD selection describes the phenomenon that the fitness of a particular genotype or phenotype is dependent on its frequency. Such context-dependency has been linked to cheating behaviour in which microbial cells that lack or have limited production of certain costly compounds benefit from other cells that do produce these compounds. When a

minimal amount of such compound increases genotype or phenotype fitness, FD selection can occur resulting in stabilization of the mutation frequency.

Early adaptations are driven by cell surface-related genes

In general, mutations that are fixed early on in the rhizosphere adaptation process tend to have a high selective advantage (32). Disruptive mutations in *oafA*, resulting in premature stops halfway the coding region, were detected as the first acquired mutations in two independent evolutionary lines (lines 1 and 3) and appear to have swept the population in the following generations (Table 2). *OafA* is part of the O-polysaccharide (O-PS, O-antigen) biosynthesis cluster 3 (*OBC3*)(29), and encodes an O-acetyltransferase which is postulated to acetylate the O-antigen component of the outer membrane LPS (33). Another *OBC3* mutation that accumulated early on in the rhizosphere adaptation process, $galE^{c.94G>A}$, leads to an amino acid substitution (V32M) in galE and this mutation swept through the population in evolutionary line 2, reaching fixation in cycle 6 (Table 2; Figure 3). *GalE* encodes an UDP-glucose 4-epimerase which is involved in O-antigen and LPS core biosynthesis (34–36). One colony with a mutation in a third *OBC3* cluster gene, *RS09880*, encoding a putative glycosyl transferase (GT) was found in cycle 6 of evolutionary line 5. Thus, in four out of the five evolutionary lines mutations that likely affect bacterial LPS structure appeared during rhizosphere adaptation, and these mutations became dominant in three out of four evolutionary lines. These results strongly suggest that modifying bacterial cell surface structure is an important bacterial strategy in early adaptation to the rhizosphere.

Adaption driven by global regulators

In the present study, six mutations were detected in *gacA* in three out of five evolutionary lines, representing approximately 20% of all missense mutations. Notably, in evolutionary lines 2 and 4, multiple *gacA* alleles accumulated, some of which were detected in early experimental cycles (Table 2). Additionally, a *gacS* mutation accumulated in line 3. *GacA* and *gacS* encode the main constituents of the conserved GacA/GacS two-component regulator system, *i.e.*, the hybrid sensor histidine kinase GacS and the cognate response regulator GacA (Figure 2). In gram-negative bacteria, activation of GacS results in cross-phosphorylation of GacA via phosphotransfer which in turn leads to activation of the expression of the small RNA genes *rsmY* and *rsmZ* via its helix-turn-helix (HTH) domain-binding domain (37). In CHA0, this regulatory pathway is known to control quorum sensing as well as secondary metabolism and stress resistance (38–41). In the closely related strain *P. protegens* Pf-5 this pathway was shown to have a big impact on bacterial genome-wide gene expression, affecting the expression of more than 10% of the annotated genes (42). Similarly, *gacA* mutants that arose on the edge of swarming Pf-5 colonies showed dramatically altered genome-wide gene expression patterns (43).

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

240 optimization of existing regulatory networks by single mutations is an important strategy for
241 bacterial adaptation to the host (48).

242

243 **Bacterial motility**

244 Bacterial motility is an important trait for rhizosphere competence, mediating colonization of
245 distal parts of the root system (49), and both LPS and the GacS/GacA two-component regulator
246 system are known to affect this trait. The O-antigen side chain of the LPS was reported to
247 contribute to swimming and swarming motility in the plant pathogenic bacterium *Erwinia*
248 *amylovora* (50, 51). The GacS/GacA two-component regulator system controls bacterial
249 motility, for example by affecting transcription of genes related to flagella and biosurfactant
250 biosynthesis (43, 49, 52). We also identified several other mutations across the various
251 evolutionary lines that can be linked to bacterial motility. For instance, a disruptive mutation in
252 the flagellar biosynthesis protein-coding gene *flhA* (H393Q.fsX15) that is involved in the
253 biogenesis of the bacterial flagellum (Table 1) appeared in evolutionary line 4, reaching up to a
254 predicted frequency of 75% (Table 2; Figure 3). In *P. fluorescens* Pf-5, and in *Pseudomonas*
255 *aeruginosa* PA01 FlhA is reported to be essential for swimming motility (42).

256 Furthermore, we identified one amino-acid substitution, R320Q, in FleQ and two in SadB,
257 R183P and L258Q respectively, that based on sequence similarity to well-studied, homologous
258 proteins in other bacteria can be linked to motility in addition to several other bacterial traits.
259 FleQ is a σ^{54}-dependent Fis family transcriptional regulator which regulates flagellar motility,
260 biofilm formation as well as Pel exopolysaccharide (EPS) production in response to cellular c-di-

GMP levels in *P. aeruginosa* (53). *P. protegens* FleQ shares 84% sequence identity and 98% sequence coverage with *P. aeruginosa* FleQ and like *P. aeruginosa* FleQ it is comprised of the N-terminal flagellar regulatory FleQ domain (PF06490), a central AAA+/ATPase σ^{54}-interaction domain (PF00158), and a C-terminal Fis-type HTH DNA-binding domain (PF02954). The R320Q substitution we identified here is found in the AAA+/ATPase σ^{54}-interacting domain in between the arginine (Arg) finger (300-303) and a c-di-GMP-binding motif, ExxxR (330-334) (53). The conserved arginine residues in FleQ, including the here mutated Arg^{320}, are thought to be important for protein oligomerization, and substitution of any of these residues abolishes ATPase activity in *Vibrio cholerae* EpsE completely (54). Finally, *sadB*, encodes a HD-related output domain (HDOD)-containing protein (Table 1) that shares 77% sequence identity and 99% sequence coverage with *P. aeruginosa* SadB. In *P. aeruginosa*, SadB stimulates Pel EPS production and the chemotaxis-like cluster CheIV, which in turn affect flagellar motility as well as biofilm formation (55). In *Pseudomonas fluorescens* F113, SadB, together with FleQ control flagellar motility, dependent and independent of the GacS/GacA two-component regulator system (56, 57).

Parallelism of targeted mutations on this functional motility pathway impelled us to assess bacterial motility and track its dynamics across all evolutionary lines. We selected all *OBC3* and *gac* mutants as well as mutants in *sadB*, *fleQ* and *flhA*. Additionally, we included two *gacA* mutant progenitors with mutations in *accC* and *RS17350*, encoding a biotin carboxylase and a methyltransferase, respectively, plus two *gacA* mutant descendants with mutations in *osmY* and *mraZ* that encode for an osmotically-inducible protein and a transcriptional repressor, respectively (Table 1). Altogether, we assessed swimming and swarming motility of seventeen

283 distinct genotypes (Figure 4; Figure S3). We found that *OBC3* mutants, the *gacA* progenitors
284 with mutations in *accC* and *RS17350*, and the *gacA* descendants with mutations in *osmY* and
285 *mraZ* were unaltered compared to their respective ancestors when considering both swimming
286 and swarming motility, with the exception of a small yet significant increase in swarming
287 motility in *oafA*^{Y335X}. However, *gac* mutants themselves were significantly affected; both in
288 swimming, which is generally enhanced, and in swarming, which is repeatedly decreased
289 (Figure 4). *SadB* mutants, like *gac* mutants display enhanced swimming and worsened
290 swarming when compared to their respective progenitors. Oddly, both the *fleQ* and the *flhA*
291 mutant displayed severely reduced swimming and swarming motility, thus representing two
292 examples of an alternative evolutionary route towards adaptation in the rhizosphere in which
293 motility is reduced. Loss of motility might coincide with another trait in these cases, such as EPS
294 production and/or biofilm formation, regulated via shared yet oftentimes opposing
295 mechanisms. Also, the frequency of the *fleQ* mutant is low, *i.e.*, only one out of six isolates from
296 cycle 6 of experimental line 5 carried this mutation and therefore whether this mutant is truly
297 beneficial is unclear. On the other hand, the *flhA* mutant was found in three out of the four
298 cultured and sequenced isolates from cycle 6 of line 4 and therefore could represent a
299 significant proportion of the total population.

300

301 **Dynamics of global phenotypic change**

302 Since natural selection eventually operates at the phenotypic level, revealing bacterial global
303 phenotypic evolutionary dynamics can help us to identify traits that are under selection.
304 Moreover, beneficial genetic mutations can be predicted if they are linked to well-known root

colonization traits. A broad range of 30 bacterial traits including different aspects of bacterial life-history traits, were examined for the sequenced isolates, which allows for genotype-phenotype association analysis.

As shown in Figure 5a, the 30 bacterial traits separated into four clusters that share a similar pattern across the different mutant genotypes and this clustering is supported by model-based clustering analysis (Figure S4). Growth of the bacteria in 1/3 strength King's B (KB) medium was positively correlated with siderophore and exoprotease production, tryptophan side oxidase activity, and growth inhibition of the bacterial plant pathogen *Ralstonia solanacearum*. Thus, this cluster, designated cluster 1, contains traits associated with bacterial social behavior, related to microbe-microbe communication and cooperation, such as the production of public goods. In a principal component analysis (Figure S5a), the first principal component (PC1) is strongly correlated with all five traits, with a total explanation of 64.5% for all variables. Cluster 2 contains traits linked to carbon source utilization. For this cluster PC1 is strongly correlated with all carbon source usage-related traits, with a total explanation of 83.9% for all variables (Figure S5b). A third cluster was observed for the bacterial ability to form a biofilm, to produce indole-3-acetic acid (IAA), and to inhibit the growth of two fungal plant pathogens, with a total explanation of 82.4% by PC1 (Figure S5c). Finally, cluster 4 contains all seven traits that are related to bacterial resistance to biotic and abiotic stresses. The first principal component (PC1) is strongly correlated with all seven traits, with a total explanation of 53.9% for all variables (Figure S5d). For these four clusters, the PC1 (or -PC1) value is used as a proxy to present the general performance of all the traits that clustered together.

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

All traits related to the utilization of 14 different carbon sources, that were selected based on their reported presence in root exudates of Arabidopsis (58) grouped in one cluster (cluster 2; Figure 5a). This suggests that carbon source utilization is co-regulated. Different mutations in *gacS* and *gacA* resulted in contrasting bacterial carbon source utilization. A mutation in the first N-terminal transmembrane domain of GacS (G27D) (Figure 2) resulted in significant enhancement of carbon source utilization (Figure 5a, c). A similar trend was observed for mutant *gacA*^{G97S}. In contrast, the majority of evolved genotypes, including most mutations in *gacA* and *rpoS*, showed a reduced ability to utilize carbon sources (Figure 5a, c).

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

348 background of the disruptive *gacA*^{E38X} mutation (Figure 5e), was the only mutational event that
349 led to a significant increase of general resistance to various environmental stresses. *FlhA*
350 encodes the flagellar biosynthesis protein FlhA and is linked to bacterial motility. The LysR-type
351 transcriptional regulator RS11785 comprises a N-terminal HTH DNA-binding domain (PF00126)
352 and a C-terminal substrate binding domain (PF03466). This substrate binding domain is linked
353 to aromatic compound degradation and resembles that of the type 2 periplasmic binding
354 proteins or PBP2s. PBP2s are responsible for binding and uptake of several substrates including
355 polysaccharides and several amino acids. We used HHpred (59) for protein remote homology
356 detection and three-dimensional structure analysis to assess the possible consequence of the
357 S256C amino acid substitution. This analysis identified OxyR from *Escherichia coli* as the best
358 template for modeling (E-value: 2.9e-30, Score: 166.45). OxyR represents the master peroxide
359 sensor in gram-negative bacteria, which operates via intramolecular disulfide-bond formation
360 in the regulatory PBP2-like domain. RS11785 is unlikely to represent the master peroxide
361 sensor in *P. protegens* CHA0 because another CHA0 protein encoded by *PFLCHA0_RS030065* is
362 much more similar to OxyR (89% percent pairwise sequence identity versus 29% for RS11785).
363 Nevertheless, it is tempting to speculate that the serine (S, Ser) to cysteine (C, Cys) amino acid
364 substitution (S256C) we observed here might influence regulatory activity by altered
365 intramolecular disulfide-bond formation as such bonds are governed by pairs of cysteine
366 residues. Lastly, RS11785 bears resemblance (37% sequence identity, 93% sequence coverage)
367 to the LysR-type PBP2-domain containing regulator alsR which regulates the activity of the
368 acetoin operon (*alsSD*) in response to several signals such as glucose and acetate in *Bacillus*
369 *amyloliquefaciens*. Acetoin (3-hydroxy-2-butanone, a precursor for 2,3-butanediol) can elicit

induced systemic resistance (60), and is linked to general plant growth promotion in *Bacilli*. In the absence of singular mutations in *flhA* and *RS11785*, the molecular mechanism underlying the observed enhanced environmental stress resistance in this double mutant remains to be clarified but we hypothesize that altered regulatory activity by RS11785^{256C} is the most likely cause.

Discussion

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

Global regulators and rhizosphere adaptation

The GacS/GacA two-component regulator system controls the production of antimicrobial secondary metabolites, exoenzymes, and siderophores, but also biofilm formation, stress responses, motility, and quorum sensing (13, 39, 42, 61, 62). In the present study, mutations in the GacS/GacA two-component regulator system caused dramatic changes in several bacterial phenotypic traits, including motility, carbon source utilization, and social traits (Figures 4 and 5). The overall gain in swimming motility is not unexpected, as swimming motility, driven by the flagellum apparatus, has been repeatedly reported to be an important root colonization trait in root-associated bacteria, including several *Pseudomonas* spp. (17, 63–65). In line with this, genome-wide transposon disruption mutant analysis by Cole and coworkers (2018) showed

395 that the majority of motility-related genes in *Pseudomonas simiae* WCS417 are positively
396 associated with Arabidopsis root colonization (23).

397 In contrast with the observed enhanced swimming motility across the evolutionary selection
398 lines, swarming motility was found to be severely hampered throughout, and this appears to be
399 the case especially for *gac* mutants. Swarming, like swimming, is driven by the flagellum but in
400 addition depends on the production of several compounds, including quorum sensing
401 molecules and biosurfactants. *P. protegens* CHA0 is a known producer of the biosurfactant
402 orfamide A and the GacS/GacA two-component system is known to be an important regulator
403 for its biosynthesis (66). These results fit observations by Song and co-workers (43) in the
404 closely related *P. protegens* strain Pf-5. During swarming motility, *gac* mutants emerged that
405 lack production of the surfactant orfamide. These mutants cannot swarm, but they co-swarm
406 with orfamide producing cells (43).

407 Remarkably, we also identified two completely non-motile mutants in our experiments with
408 disruptions in the *fleQ* and *flhA* genes. Possibly, these mutants adhere much better to the root
409 surface or might be able to form a biofilm more rapidly or strongly as these traits are often
410 inversely correlated with bacterial motility (67–69). Future experiments should reveal whether
411 such trade-off between bacterial motility and root adherence indeed underlies our
412 observations in this evolutionary experiment.

413 The decreased production of public goods, such as siderophores and exoproteases, which were
414 observed in all independent evolutionary lines (Figure 5), could be beneficial for bacterial
415 fitness by saving energy and primary metabolites. For example, adaptation of *Pseudomonas*

aeruginosa to the human host through mutations in regulators was accompanied by loss of siderophore production, secreted proteases, and biofilm formation (70, 71).

Cell surface structures as bacterial adaptation targets

Bacterial cell surface components are the first line of defense against environmental stress and interplay with hosts (72–74). LPS is a central outer membrane component for gram-negative bacteria, and exhibits structural adaptability that is contributed especially by its O-antigen part (72, 74). Bacterial O-antigen structure modification plays an important role in evasion of host immunity (75), and has the potential to change host-bacteria interactions (72, 74, 76). In plant pathogenic bacteria, LPS components are important virulence determinants (77, 78), which can activate a variety of defense-related responses (77, 79–81). In *P. fluorescens* the O-antigen component has been implicated to induce systemic resistance in radish (82) and Arabidopsis (83).

We observed parallel mutations in genes that are involved in LPS biosynthesis and structure modification. In three out of the five evolutionary lines, the first fixed mutations were identified in *oafA* and *galE* that are annotated as O-antigen biosynthesis and structure modification related (Table 1). *OafA* encodes an O-acetyltransferase, which is postulated to modify the O-antigen by acetylation (33). The enzyme GalE, UDP-galactose 4-epimerase, is involved in the interconversion of UDP-glucose to UDP-galactose, an essential intermediate for LPS core and O-antigen structures (34–36). Inactivation of *galE* in *Porphyromonas gingivalis* resulted in shortening of the O-antigen (84), and in *Bradyrhizobium japonicum*, disruption of *galE* resulted

437 in the complete absence of O-antigen (34). Thus, it is tempting to speculate that evasion of the
438 plant's immune response plays a role in adaptation of CHA0 to the rhizosphere of Arabidopsis.

439 In conclusion, the observed bacterial genetic and phenotypic adaption dynamics emphasize
440 important roles for global regulators, motility, and cell surface structure in bacterial adaptation
441 to its host. The parallel emergence of mutations in similar genes resulted in specific fitness
442 advantages for mutants in the rhizosphere, suggesting that this evolutionary process is driven
443 by the rhizosphere environment.

444

445 ## Materials and Methods

446 ### Experimental setup

447 We set up an experimental evolution experiment with *Arabidopsis thaliana* (Arabidopsis)
448 ecotype Col-0 as host plant and *Pseudomonas protegens* CHA0 (CHA0) as the evolving bacterial
449 strain. CHA0 (85) is a model strain originally isolated from roots of tobacco plants grown in soil
450 naturally suppressive to black root rot (86). CHA0 was chromosomally tagged with GFP and a
451 kanamycin resistance cassette (Jousset et al 2006) to enable consistent tracking of the strain
452 and identification of contaminations. We previously described the setup of the evolutionary
453 experiment in great detail (26). In brief: the ancestral bacterial population (10^6 cells) was
454 inoculated on Arabidopsis roots grown under gnotobiotic conditions inside ECO2 boxes in
455 carbon-free silver sand. For each cycle, Arabidopsis seeds were surface sterilized using chlorine
456 gas, germinated on modified Hoagland's agar medium, and grown for two weeks until
457 transplantation into the ECO2 boxes, containing each two plants (26). Inoculated two-week-old

458 seedlings were then grown for an additional four weeks, after which the root-associated
459 bacteria were collected in 10 mM $MgSO_4$ and subjected to fluorescence-based cell counting by
460 flow cytometry yielding on average 10^7 cells/root. 10^6 cells of the evolved bacterial populations
461 were then transferred to new plants and this cycle was repeated eight times. After each cycle, a
462 small fraction of each population was plated on general-purpose, nonselective medium, 3 g/l
463 tryptic soy agar (TSA), to assess for contaminations and to verify that all colonies carried the
464 *GFP* marker gene, as observed under UV light, on the one hand, and to select individual isolates
465 for phenotypic characterization on the other hand.

466 We previously picked sixteen random isolates from each of five experimental lines at cycles 2, 4
467 and 6 respectively plus sixteen isolates from the ancestor population, yielding a total set of 256
468 isolates (26). These 256 isolates were subsequently grouped into five distinct phenotypes based
469 on their performance on a variety of bacterial life-history traits (26). In the current study we
470 selected six of these sixteen isolates per line, and per cycle, for genome sequence analysis.
471 Isolates were selected to represent the breath of phenotypic diversity observed previously (26).
472 Together with six isolates from the ancestor population we set out to obtain genome sequences
473 of 96 isolates in total using the NextSeq-500 Illumina platform (2 x 75 bp paired-end).
474 Sequencing of two isolates from line 4, cycle 4, however, failed and thus a final set of 94
475 genomes were retrieved. We then used the snippy pipeline
476 (https://github.com/tseemann/snippy), integrating reference genome-based mapping of the
477 Illumina reads by BWA-MEM, variant calling by SAMtools and FreeBayes, and variant impact
478 analysis using SnpEff (87), to identify single nucleotide polymorphisms (SNPs) and small indels
479 and deletions (INDELs). Larger INDELs were identified by calculating the breath of coverage of

480 the mapped Illumina reads on the reference genome in a sliding window using bedtools (88).
481 Regions with reduced coverage (<99%) were manually inspected in the Integrative Genome
482 Viewer (IGV). Phylogenetic trees for each line were constructed manually with illustrator based
483 on all detected mutations, the length of the branches representing the number of mutations.
484 The genealogy and frequency of each lineage is shown in the Muller plots that are prepared
485 with R package '*ggmuller*'.

486

487 **Bacterial life-history traits**

488 For the 94 sequenced isolates, a variety of bacterial life-history traits reflecting various aspects
489 of bacterial physiological processes, were measured previously as part of all 256 isolates initially
490 collected (26). Briefly, we monitored optical density (OD) at a wavelength of 600 nm to
491 estimate the bacterial yield after 72 hours of growth under different growth conditions in 96-
492 well microplates. We measured bacterial growth yield and resistance to various stresses,
493 including acidic (pH = 5) and alkaline (pH = 9) conditions, oxidative stress in 0.0025% H_2O_2,
494 water potential stress (15% polyethylene glycol (PEG)-6000)), and salt stress (2% NaCl), and
495 resistance to the antibiotics streptomycin (1 $\mu g \cdot ml^{-1}$), tetracycline (1 $\mu g \cdot ml^{-1}$), and penicillin (5
496 $\mu g \cdot ml^{-1}$). Bacterial carbon source utilization was quantified as growth yield in modified Ornston
497 and Stanier (OS) minimal medium (89) supplemented with single carbon sources that have been
498 reported to be abundant in Arabidopsis root exudates (Chaparro et al 2013). These included the
499 following carbon sources; alanine, arabinose, butyrolactam, fructose, galactose, glucose,
500 glycerol, glycine, lactic acid, putrescine, serine, succinic acid, threonine, and valine that were
501 added to a final concentration of 0.5 $g \cdot l^{-1}$. In addition, we measured bacterial auxin (indole-3-

acetic acid or IAA) production with a colorimetric test (90), iron-chelating ability using a modified chrome azurol S (CAS) assay (91), proteolytic activity by the adapted assay from Smeltzer et al. (1993), tryptophan side chain oxidase activity using a colorimetric assay (92), and biofilm formation using a modified Chrystal Violet staining assay (93). We measured the OD values reflecting the color intensities at specific wavelengths to quantify these traits. We further assessed bacterial antimicrobial activity by quantifying their effect on growth of the fungi *Verticillium dahliae* and *Fusarium oxysporum,* and the bacterium *Ralstonia solanacearum*.

Motility assays

Motility assays were undertaken in round petri dish plates containing Cook's Cytophaga medium (CC medium) (0.3% agar for swimming, 0.5% agar for swarming) (94), using typical swimming and swarming assays as described by Déziel et al (95). All tested strains were grown on King's B medium agar plates for 24 hours before inoculation. Swim and swarm plates were inoculated with the tested strains with a sterile toothpick. For swimming plates, the inoculum was introduced by gently piercing the agar such that the motility within the semisolid agar could be evaluated. For swarming plates, the inoculum was introduced on the agar surface enabling visualization of motility across the agar surface. Both swimming and swarming plates were imaged after 18 hours incubation at 21 °C with right-side-up. The radii of swimming and swarming motility were determined from the photographs by ImageJ, examining the circular turbid zone inside the agar for swimming and the circular zone on the agar surface for swarming.

523

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

525 *Hierarchical clustering*

526 A heatmap to illustrate the association patterns of bacterial genotypes and their measured
527 traits was constructed in R using the *ggplot2* package. Isolates without mutations or only with
528 synonymous mutations were excluded from the association analysis. Hierarchical clustering was
529 performed using the Ward.D2 method, that is based on the squared dissimilarities of the set of
530 objects being clustered (96).

531 *Model-based clustering*

532 We applied a model-based clustering method to reveal the best fitting structures of trait
533 covariance patterns. For example, some traits might be either directly or indirectly co-regulated
534 by the same gene, which is expected for global regulators particularly which can co-regulate
535 thousands of genes. We used the *mclust* package in R to run the model simulation (97). This
536 method assumes that the input data are distributed as a mixture of two or more clusters. The
537 advantage of the model-based clustering method is that it avoids heuristic assumptions and
538 uses a soft assignment that every data point has a possibility of falling to each cluster, which
539 facilitates the best clustering solution. The so-called Bayesian Information Criterion (BIC) was
540 used to select the best model. A large BIC score indicates a better fit of the model.

541 This result is in line with the outcome of hierarchical clustering with Adjusted Rand Index (ARI)
542 set as 1, and *k* set as 4 in "ward.D2" as indicated in Figure 5a. ARI is usually used to evaluate the
543 match degree of a given clustering solution comparing to the model-based clustering result,

544 with 0 reflecting a random partition and 1 the boundary of accuracy of a certain clustering
545 solution (97).

546

547 **Genotype-phenotype association analysis**

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

555 To examine the accumulative effects of each mutation on bacterial phenotype, ANOVA was
556 used to compare cluster proxies of evolved genotypes with their direct ancestors. Only
557 genotypes identified more than once (n>=2) were included in this analysis.

558

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

560 We used High-Resolution Melting (HRM) profile analysis with integrated LunaProbes to quantify
561 the ratio of mutant to wild type genotypes (98–100). The probes and primers used in this study
562 are listed in Table S2. Primers were designed using Primer3. Probes were designed with the
563 single nucleotide polymorphism (SNP) located in the middle of the sequence, and the 3' end
564 was blocked by carbon spacer C3. The primer asymmetry was set to 2:1 (excess primer: limiting

565 primer) in all cases. Pre-PCR was performed in a 10-µl reaction system, with 0.25 µM excess
566 primer, 0.125 µM limiting primer, 0.25 µM probe, 0.5 µl bacterial sample culture (100-fold
567 diluted saved sample, OD_{600} is about 0.01), 1X LightScanner Master Mix (BioFire Defense).
568 DMSO with the final concentration 5% was supplemented in all reactions to ensure the
569 targeted melting domains are within the detection limit of the LightScanner (Idaho Technology
570 Inc.). Finally, MQ water was used to supplement up to 10 µl. A 96-well black microtiter plate
571 with white wells was used to minimize background fluorescence. Before amplification, 25 µl
572 mineral oil was loaded in each well to prevent evaporation, and the plate was covered with a
573 foil seal to prevent the degradation of fluorescent molecules. Amplification was initiated by a
574 holding at 95 °C for 3 min, followed by 55 cycles of denaturation at 95 °C for 30 s, annealing at
575 60 °C for 30 s and extension at 72 °C for 30 s and then kept at 72 °C for 10 min. After
576 amplification, samples were heated in a ThermalCycler (Bio-Rad) shortly to 95 °C for 30 s to
577 denature all double-stranded structures followed by a rapid cooling to 25 °C for 30 s to facilitate
578 successful hybridization between probes and the target strands. The plate was then transferred
579 to a LightScanner (Idaho Technology Inc.). Melting profiles of each well were collected by
580 monitoring the continuous loss of fluorescence with a steady increase of the temperature from
581 35 °C to 97 °C with a ramp rate of 0.1 °C /s. The relative quantification was based on the
582 negative first derivative plots using software MATLAB. The areas of probe-target duplexes
583 melting peaks were auto-calculated by 'AutoFit Peaks I Residuals' function in software PeakFit
584 (SeaSolve Software Inc.). The mutant frequency *X* was calculated using the formula shown
585 below:

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

586 To validate the HRM method, standard curves were generated by measuring mixed samples
587 with known proportions of mutant templates: 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%,
588 90% and 100%. Measurements for each sample were done in triplicate. Linear regression
589 formula of each mutant between actual frequencies and measured frequencies were shown in
590 Figure S1. The high R^2 values, and nearly equal to 1 slope values of these equations, confirmed
591 that the HRM method can accurately detect mutants' frequency in a mixed population.

592

593 ## Data availability

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

598

599 ## Acknowledgments

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

603

Author contributions

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

References

1. Bakker PAHM, Berendsen RL, Van Pelt JA, Vismans G, Yu K, Li E, Van Bentum S, Poppeliers SWM, Sanchez Gil JJ, Zhang H, Goossens P, Stringlis IA, Song Y, de Jonge R, Pieterse CMJ. 2020. The soil-borne identity and microbiome-assisted agriculture: looking back to the future. Mol Plant 13:1394–1401.
2. Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM. 2006. The role of root exudates in rhizosphere interations with plants and other organisms. Annu Rev Plant Biol 57:233–266.
3. Stringlis IA, Yu K, Feussner K, de Jonge R, Van Bentum S, Van Verk MC, Berendsen RL, Bakker PAHM, Feussner I, Pieterse CMJ. 2018. MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. Proc Natl Acad Sci U S A 115:E5213–E5222.
4. Bakker PAHM, Pieterse CMJ, de Jonge R, Berendsen RL. 2018. The soil-borne legacy. Cell 172:1178–1180.
5. Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH. 2013. Going back to the roots: the microbial ecology of the rhizosphere. Nat Rev Microbiol 11:789–799.

6. Sasse J, Martinoia E, Northen T. 2018. Feed your friends: do plant exudates shape the root microbiome? Trends Plant Sci 23:25–41.

7. Bais HP, Park SW, Weir TL, Callaway RM, Vivanco JM. 2004. How plants communicate using the underground information superhighway. Trends Plant Sci 9:26–32.

8. Zamioudis C, Pieterse CMJ. 2012. Modulation of host immunity by beneficial microbes. Mol Plant-Microbe Interact 25:139–150.

9. Yang J, Kloepper JW, Ryu C-M. 2009. Rhizosphere bacteria help plants tolerate abiotic stress. Trends Plant Sci 14:1–4.

10. Browne P, Rice O, Miller SH, Burke J, Dowling DN, Morrissey JP, O'Gara F. 2009. Superior inorganic phosphate solubilization is linked to phylogeny within the *Pseudomonas fluorescens* complex. Appl Soil Ecol 43:131–138.

11. Lugtenberg B, Kamilova F. 2009. Plant growth promoting rhizobacteria. Annu Rev Microbiol 63:541–556.

12. Ryu C-M, Farag MA, Hu C-H, Reddy MS, Wei H-X, Paré PW, Kloepper JW. 2003. Bacterial volatiles promote growth in *Arabidopsis*. Proc Natl Acad Sci U S A 100:4927–4932.

13. Haas D, Défago G. 2005. Biological control of soil-borne pathogens by fluorescent pseudomonads. Nat Rev Microbiol 3:307–319.

14. Pieterse CMJ, Zamioudis C, Berendsen RL, Weller DM, Van Wees SCM, Bakker PAHM. 2014. Induced systemic resistance by beneficial microbes. Annu Rev Phytopathol 52:347–375.

15. Raaijmakers JM, Bonsall RF, Weller DM. 1999. Effect of population density of *Pseudomonas fluorescens* on production of 2,4-Diacetylphloroglucinol in the rhizosphere

of wheat. Phytopathology 89:470–475.

16. Raaijmakers JM, Leeman M, Oorschot van MM., et.al. 1995. Dose-response relationships in biological control of Fusarium wilt of radish by *Pseudomonas* spp.. Phytopathology 85:1075–1081.

17. Lugtenberg BJJ, Dekkers L, Bloemberg G V. 2001. Molecular determinants of rhizosphere colonization *Pseudomonas*. Annu Rev Phytopathol 39:461–490.

18. Pieterse CMJ, Berendsen RL, de Jonge R, Stringlis IA, Van Dijken AJH, Van Pelt JA, Van Wees SCM, Yu K, Zamioudis C, Bakker PAHM. 2020. *Pseudomonas simiae WCS417*: star track of a model beneficial rhizobacterium. Plant Soil. Springer Science and Business Media Deutschland GmbH.

19. de Weger LA, van der Vlugt CI, Wijfjes AH, Bakker PA, Schippers B, Lugtenberg B. 1987. Flagella of a plant-growth-stimulating *Pseudomonas fluorescens* strain are required for colonization of potato roots. J Bacteriol 169:2769–2773.

20. de Weger LA, van Loosdrecht MC, Klaassen HE, Lugtenberg B. 1989. Mutational changes in physiochemical cell surface properties of plant-growth-stimulating *Pseudomonas* spp. do not influence the attachment properties of the cells. J Bacteriol 171:2756–2761.

21. Simons M, Permentier HP, de Weger LA, Wijffelman CA, Lugtenberg BJJ. 1997. Amino acid synthesis is necessary for tomato root colonization by *Pseudomonas fluorescens* strain WCS365. Mol Plant-Microbe Interact 10:102–106.

22. Yu K, Liu Y, Tichelaar R, Savant N, Lagendijk E, van Kuijk SJL, Stringlis IA, van Dijken AJH, Pieterse CMJ, Bakker PAHM, Haney CH, Berendsen RL. 2019. Rhizosphere-associated *Pseudomonas* suppress local root immune responses by gluconic acid-mediated lowering

669 of environmental pH. Curr Biol 29:3913-3920.e4.

670 23. Cole BJ, Feltcher ME, Waters RJ, Wetmore KM, Mucyn TS, Ryan EM, Wang G, Ul-Hasan S,

671 McDonald M, Yoshikuni Y, Malmstrom RR, Deutschbauer AM, Dangl JL, Visel A. 2017.

672 Genome-wide identification of bacterial plant colonization genes. PLoS Biol 15:e2002860.

673 24. Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC. 2012. Experimental

674 evolution. Trends Ecol Evol 27:547–560.

675 25. Jousset A, Schuldes J, Keel C, Maurhofer M, Daniel R, Scheu S, Thuermer A. 2014. Full-

676 genome sequence of the plant growth-promoting bacterium *Pseudomonas protegens*

677 CHA0. Genome Announc 2:e00322-14.

678 26. Li E, de Jonge R, Liu C, Jiang H, Friman V-P, AHM Bakker P, Jousset A. 2020. Rapid

679 evolution of bacterial mutualism in the plant rhizosphere. bioRxiv

680 https://doi.org/10.1101/2020.12.07.414607.

681 27. Oren Y, Smith MB, Johns NI, Kaplan Zeevi M, Biran D, Ron EZ, Corander J, Wang HH, Alm

682 EJ, Pupko T. 2014. Transfer of noncoding DNA drives regulatory rewiring in bacteria. Proc

683 Natl Acad Sci U S A 111:16112–16117.

684 28. Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of

685 morphological evolution. Cell 134:25–36.

686 29. Kupferschmied P, Chai T, Flury P, Blom J, Smits THM, Maurhofer M, Keel C. 2016. Specific

687 surface glycan decorations enable antimicrobial peptide resistance in plant-beneficial

688 *pseudomonads* with insect-pathogenic properties. Environ Microbiol 18:4265–4281.

689 30. Galperin MY, Makarova KS, Wolf YI, Koonin E V. 2015. Expanded Microbial genome

690 coverage and improved protein family annotation in the COG database. Nucleic Acids Res

691 43:D261–D269.

692 31. Maddamsetti R, Lenski RE, Barrick JE. 2015. Adaptation, clonal interference, and
693 frequency-dependent interactions in a long-term evolution experiment with escherichia
694 coli. Genetics 200:619–631.

695 32. Barrett RDH, Maclean RC, Bell G. 2006. Mutations of intermediate effect are responsible
696 for adaptation in evolving *Pseudomonas fluorescens* populations. Biol Lett 2:236–238.

697 33. Allison GE, Verma NK. 2000. Serotype-converting bacteriophages and O-antigen
698 modification in *Shigella flexneri*. Trends Microbiol 8:17–23.

699 34. Chang W-S, Park K-M, Koh S-C, So J-S. 2008. Characterization of the *Bradyrhizobium*
700 *japonicum galE* gene: its impact on lipopolysaccharide profile and nodulation of soybean.
701 FEMS Microbiol Lett 280:242–249.

702 35. Liu B, Knirel YA, Feng L, Perepelov A V., Senchenkova SN, Wang Q, Reeves PR, Wang L.
703 2008. Structure and genetics of *Shigella* O antigens. FEMS Microbiol Rev 32:627–653.

704 36. Samuel G, Reeves P. 2003. Biosynthesis of O-antigens: genes and pathways involved in
705 nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydr Res 338:2503–
706 2519.

707 37. Brencic A, McFarland KA, McManus HR, Castang S, Mogno I, Dove SL, Lory S. 2009. The
708 GacS/GacA signal transduction system of *Pseudomonas aeruginosa* acts exclusively
709 through its control over the transcription of the RsmY and RsmZ regulatory small RNAs.
710 Mol Microbiol 73:434–445.

711 38. Brown D. 2010. A mathematical model of the Gac/Rsm quorum sensing network in
712 *Pseudomonas fluorescens*. BioSystems 101:200–212.

713 39. Heeb S, Blumer C, Haas D. 2002. Regulatory RNA as mediator in GacA/RsmA-dependent
714 global control of exoproduct formation in *Pseudomonas fluorescens* CHA0. J Bacteriol
715 184:1046–1056.
716 40. Valverde C, Heeb S, Keel C, Haas D. 2003. RsmY, a small regulatory RNA , is required in
717 concert with RsmZ for GacA-dependent expression of biocontrol traits in *Pseudomonas*
718 *fluorescens* CHA0. Mol Microbiol 50:1361–1379.
719 41. Heeb S, Valverde C, Haas D. 2005. Role of the stress sigma factor RpoS in GacA / RsmA-
720 controlled secondary metabolism and resistance to oxidative stress in *Pseudomonas*
721 *fluorescens* CHA0. FEMS Microbiol Lett 243:251–258.
722 42. Hassan KA, Johnson A, Shaffer BT, Ren Q, Kidarsa TA, Elbourne LDH, Hartney S, Duboy R,
723 Goebel NC, Zabriskie TM, Paulsen IT, Loper JE. 2010. Inactivation of the GacA response
724 regulator in *Pseudomonas fluorescens* Pf-5 has far-reaching transcriptomic
725 consequences. Environ Microbiol 12:899–915.
726 43. Song C, Kidarsa TA, Mortel JE Van De, Loper JE, Jos M. 2016. Living on the edge:
727 emergence of spontaneous gac mutations in *Pseudomonas protegens* during swarming
728 motility. Environ Microbiol 18:3453–3465.
729 44. Hunter P. 2008. The great leap forward. EMBO Rep 9:608–611.
730 45. Zambrano MM, Siegele DA, Almirón M, Tormo A, Kolter R. 1993. Microbial competition:
731 *Escherichia coli* mutants that take over stationary phase cultures. Science 259:1757–
732 1760.
733 46. Sumby P, Whitney AR, Graviss EA, DeLeo FR, Musser JM. 2006. Genome-wide analysis of
734 group A streptococci reveals a mutation that modulates global phenotype and disease

735 specificity. PLoS Pathog 2:e5.

736 47. Giraud A, Arous S, De Paepe M, Gaboriau-Routhiau V, Bambou JC, Rakotobe S, Lindner
737 AB, Taddei F, Cerf-Bensussan N. 2008. Dissecting the genetic components of adaptation
738 of *Escherichia coli* to the mouse gut. PLoS Genet 4:0052–0061.

739 48. Damkiaer S, Yang L, Molin S, Jelsbak L. 2013. Evolutionary remodeling of global
740 regulatory networks during long-term bacterial adaptation to human hosts. Proc Natl
741 Acad Sci U S A 110:7766–7771.

742 49. Sánchez-Contreras M, Martín M, Villacieros M, O'Gara F, Bonilla I, Rivilla R. 2002.
743 Phenotypic selection and phase variation occur during alfalfa root colonization by
744 *Pseudomonas fluorescens* F113. J Bacteriol 184:1587–1596.

745 50. Bowden SD, Hale N, Chung JCS, Hodgkinson JT, Spring DR, Welch M. 2013. Surface
746 swarming motility by *Pectobacterium atrosepticum* is a latent phenotype that requires O
747 antigen and is regulated by quorum sensing. Microbiology 159:2375–2385.

748 51. Berry MC, McGhee GC, Zhao Y, Sundin GW. 2009. Effect of a *waaL* mutation on
749 lipopolysaccharide composition, oxidative stress survival, and virulence in *Erwinia*
750 *amylovora*. FEMS Microbiol Lett 291:80–87.

751 52. Lapouge K, Schubert M, Allain FH-T, Haas D. 2007. Gac/Rsm signal transduction pathway
752 of γ-proteobacteria: from RNA recognition to regulation of social behaviour. Mol
753 Microbiol 67:241–253.

754 53. Matsuyama BY, Krasteva P V., Baraquet C, Harwood CS, Sondermann H, Navarro MVAS.
755 2016. Mechanistic insights into c-di-GMP-dependent control of the biofilm regulator FleQ
756 from *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A 113:E209–E218.

757 54. Patrick M, Korotkov K V., Hol WGJ, Sandkvist M. 2011. Oligomerization of EpsE
758 coordinates residues from multiple subunits to facilitate ATPase activity. J Biol Chem
759 286:10378–10386.
760 55. Caiazza NC, O'Toole GA. 2004. SadB is required for the transition from reversible to
761 irreversible attachment during biofilm formation by *Pseudomonas aeruginosa* PA14. J
762 Bacteriol 186:4476–4485.
763 56. Muriel C, Blanco-Romero E, Trampari E, Arrebola E, Durán D, Redondo-Nieto M, Malone
764 JG, Martín M, Rivilla R. 2019. The diguanylate cyclase AdrA regulates flagellar
765 biosynthesis in *Pseudomonas fluorescens* F113 through SadB. Sci Rep 9:1–9.
766 57. Martínez-Granero F, Navazo A, Barahona E, Redondo-Nieto M, Rivilla R, Martín M. 2012.
767 The Gac-Rsm and SadB signal transduction pathways converge on AlgU to downregulate
768 motility in *Pseudomonas fluorescens*. PLoS One 7:e31765.
769 58. Chaparro JM, Badri D V., Bakker MG, Sugiyama A, Manter DK, Vivanco JM. 2013. Root
770 exudation of phytochemicals in *Arabidopsis* follows specific patterns that are
771 developmentally programmed and correlate with soil microbial functions. PLoS One
772 8:e55731.
773 59. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas
774 AN, Alva V. 2018. A completely reimplemented MPI bioinformatics toolkit with a new
775 HHpred server at its core. J Mol Biol 430:2237–2243.
776 60. Rudrappa T, Biedrzycki ML, Kunjeti SG, Donofrio NM, Czymmek KJ, Paré PW, Bais HP.
777 2010. The rhizobacterial elicitor acetoin induces systemic resistance in *Arabidopsis*
778 *thaliana*. Commun Integr Biol 3:130–138.

779 61. Haas D, Keel C. 2003. Regulation of antibotic production in root-colonizing *Pseudomonas*
780 spp. and relavance for biological control of plant disease. Annu Rev Phytopathol 41:117–
781 153.
782 62. Raaijmakers JM, De Bruijn I, Nybroe O, Ongena M. 2010. Natural functions of
783 lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics.
784 FEMS Microbiol Rev 34:1037–1062.
785 63. Knights HE, Jorrin B, Haskett TL, Poole PS. 2021. Deciphering bacterial mechanisms of
786 root colonization. Environ Microbiol Rep https://doi.org/10.1111/1758-2229.12934.
787 64. Colaianni NR, Parys K, Lee H-S, Conway JM, Kim NH, Edelbacher N, Mucyn TS, Madalinski
788 M, Law TF, Jones CD, Belkhadir Y, Dangl JL. 2021. A complex immune response to flagellin
789 epitope variation in commensal communities. Cell Host Microbe
790 https://doi.org/10.1016/j.chom.2021.02.006.
791 65. Parys K, Colaianni NR, Lee H-S, Hohmann U, Edelbacher N, Trgovcevic A, Blahovska Z, Lee
792 D, Mechtler A, Muhari-Portik Z, Madalinski M, Schandry N, Rodríguez-Arévalo I, Becker C,
793 Sonnleitner E, Korte A, Bläsi U, Geldner N, Hothorn M, Jones CD, Dangl JL, Belkhadir Y.
794 2021. Signatures of antagonistic pleiotropy in a bacterial flagellin epitope. Cell Host
795 Microbe https://doi.org/10.1016/j.chom.2021.02.008.
796 66. Sobrero PM, Muzlera A, Frescura J, Jofré E, Valverde C. 2017. A matter of hierarchy:
797 activation of orfamide production by the post-transcriptional Gac-Rsm cascade of P
798 seudomonas protegens CHA0 through expression upregulation of the two dedicated
799 transcriptional regulators. Environ Microbiol Rep 9:599–611.
800 67. Romling U, Galperin MY, Gomelsky M. 2013. Cyclic di-GMP: the first 25 years of a

universal bacterial second messenger. Microbiol Mol Biol Rev 77:1–52.

68. Caiazza NC, Merritt JH, Brothers KM, O'Toole GA. 2007. Inverse regulation of biofilm formation and swarming motility by *Pseudomonas aeruginosa* PA14. J Bacteriol 189:3603–3612.

69. Simm R, Morr M, Kader A, Nimtz M, Römling U. 2004. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessibility to motility. Mol Microbiol 53:1123–1134.

70. Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, D'Argenio DA, Miller SI, Ramsey BW, Speert DP, Moskowitz SM, Burns JL, Kaul R, Olson M V. 2006. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. Proc Natl Acad Sci U S A 103:8487–92.

71. Winstanley C, O'Brien S, Brockhurst MA. 2016. *Pseudomonas aeruginosa* evolutionary adaptation and diversification in cystic fibrosis chronic lung infections. Trends Microbiol 24:327–337.

72. Lerouge I, Vanderleyden J. 2002. O-antigen structural variation: mechanisms and possible roles in animal/plant–microbe interactions. FEMS Microbiol Rev 26:17–47.

73. Whitfield C, Trent MS. 2014. Biosynthesis and export of bacterial lipopolysaccharides. Annu Rev Biochem 83:99–128.

74. Maldonado RF, Sá-Correia I, Valvano MA. 2016. Lipopolysaccharide modification in Gram-negative bacteria during chronic infection. FEMS Microbiol Rev 40:480–493.

75. Murray GL, Attridge SR, Morona R. 2006. Altering the length of the lipopolysaccharide O antigen has an impact on the interaction of *Salmonella enterica* serovar Typhimurium

with macrophages and complement. J Bacteriol 188:2735–2739.

76. Ranf S. 2016. Immune sensing of lipopolysaccharide in plants and animals: same but different. PLoS Pathog 12:e1005596.

77. Newman M-A, von Roepenack E, Daniels M, Dow M. 2000. Lipopolysaccharides and plant responses to phytopathogenic bacteria. Mol Plant Pathol 1:25–31.

78. Di Lorenzo F, Silipo A, Andersen Gersby LB, Palmigiano A, Lanzetta R, Garozzo D, Boyer C, Pruvost O, Newman MA, Molinaro A. 2017. *Xanthomonas citri* pv. *citri* pathotypes: LPS structure and function as microbe-associated molecular patterns. ChemBioChem 18:772–781.

79. Newman MA, Daniels MJ, Dow JM. 1995. Lipopolysaccharide from *Xanthomonas campestris* induces defense-related gene expression in *Brassica campestris*. Mol Plant-Microbe Interact 8:778–80.

80. Barton-Willis PA, Wang MC, Holliday MJ, Long MR, Keen NT. 1984. Purification and composition of lipopolysaccharides from *Pseudomonas syringae*. pv. *glycinea*. Physiol Plant Pathol 25:387–398.

81. Erbs G, Newman M-A. 2003. The role of lipopolysaccharides in induction of plant defence responses. Mol Plant Pathol 4:421–425.

82. Leeman M, Pelt JA van, Ouden FM den, Heinsbroek M, Bakker PAHM, Schippers B. 1995. Induction of systemic resistance against *Fusarium wilt* of radish by lipopolysaccharides of *Pseudomonas fluorescens*. Phytopathol 85:1021–1027.

83. Van Wees SCM, Pieterse CMJ, Trijssenaar A, Van 't Westende YAM, Hartog F, Van Loon LC. 1997. Differential induction of systemic resistance in *Arabidopsis* by biocontrol

bacteria. Mol Plant-Microbe Interact 10:716–724.

84. Nakao R, Senpuku H, Watanabe H. 2006. *Porphyromonas gingivalis galE* is involved in lipopolysaccharide O-antigen synthesis and biofilm formation. Infect Immun 74:6145–53.

85. Ramette A, Frapolli M, Saux MF-L, Gruffaz C, Meyer J-M, Défago G, Sutra L, Moënne-Loccoz Y. 2011. *Pseudomonas protegens* sp. nov., widespread plant-protecting bacteria producing the biocontrol compounds 2,4-diacetylphloroglucinol and pyoluteorin. Syst Appl Microbiol 34:180–188.

86. Stutz EW, Défago G, Kern H. 1986. Naturally occurring fluorescent pseudomonads involved in suppression of black root rot of tobacco. Phytopathology 76:181–185.

87. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w1118; iso-2; iso-3*. Fly (Austin) 6:80–92.

88. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

89. Højberg O, Schnider U, Winteler H V, Sørensen J, Haas D. 1999. Oxygen-sensing reporter strain of *Pseudomonas fluorescens* for monitoring the distribution of low-oxygen habitats in soil. Appl Environ Microbiol 65:4085–4093.

90. Glickmann E, Dessaux Y. 1995. A critical examination of the specificity of the salkowski reagent for indolic compounds produced by phytopathogenic bacteria. Appl Environ Microbiol 61:793–796.

91. Alexander DB, Zuberer DA. 1991. Use of chrome azurol S reagents to evaluate

867 siderophore production by rhizosphere bacteria. Biol Fertil Soils 12:39–45.

868 92. Oberhänsli T, Défago G, Haas D. 1991. Indole-3-acetic acid (IAA) synthesis in the
869 biocontrol strain CHA0 of *Pseudomonas fluorescens*: role of tryptophan side chain
870 oxidase. Microbiology 137:2273–2279.

871 93. Moskowitz SM, Foster JM, Emerson J, Burns JL. 2004. Clinically feasible biofilm
872 susceptibility assay for isolates of *Pseudomonas aeruginosa* from patients with cystic
873 fibrosis. J Clin Microbiol 42:1915–1922.

874 94. Christensen PJ, Cook FD. 1972. The isolation and enumeration of cytophagas. Can J
875 Microbiol 18:1933–1940.

876 95. Déziel E, Comeau Y, Villemur R. 2001. Initiation of biofilm formation by *Pseudomonas*
877 *aeruginosa* 57RP correlates with emergence of hyperpiliated and highly adherent
878 phenotypic variants deficient in swimming, swarming, and twitching motilities. J Bacteriol
879 183:1195–1204.

880 96. Murtagh F, Legendre P. 2014. Ward's hierarchical agglomerative clustering method:
881 which algorithms implement ward's criterion? J Classif 31:274–295.

882 97. Scrucca L, Fop M, Murphy TB, Raftery AE. 2016. Mclust 5: clustering, classification and
883 density estimation using gaussian finite mixture models. R J 8:289–317.

884 98. Kai Lee H, Kiat Lee C, Ping Loh T, Wei-Tze Tang J, Anantharajah Tambyah P, Siew-Chuan
885 Koay E. 2011. High-resolution melting approach to efficient identification and
886 quantification of H275Y mutant influenza H1N1/2009 virus in mixed-virus-population
887 samples. J Clin Microbiol 49:3555–3559.

888 99. Zhong W-L, Wang L, Wu X, Zhang J, Chen X-F, Zhang W, Dou X, Yu B. 2016. Development

889 of unlabeled probe based high-resolution melting analysis for detection of filaggrin gene
890 mutation c.3321delA. J Clin Lab Anal 30:892–896.
891 100. Capper RL, Jin YK, Lundgren PB, Peplow LM, Matz M V., van Oppen MJH. 2015.
892 Quantitative high resolution melting: two methods to determine SNP allele frequencies
893 from pooled samples. BMC Genet 16:62.
894 101. Noble R. 2019. Create Muller Plots of Evolutionary Dynamics [R package ggmuller version
895 0.5.4].
896
897

898 **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

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

900 **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	
						9.2%	19.1%	46.7%	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	
						5.5%	4.5%	15.5%	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	

						3.8%	19.9%	75.6%	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	

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

[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

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

Figure legends

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

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).

927 **Figure 3**. Phylogenetic trees and Muller plots of five independently evolving *P. protegens* CHA0
928 populations. Left) Phylogenies for 18 genomes from each population (16 for the population in
929 experimental line 4), based on the sequential appearance of mutations are shown. Synonymous
930 mutations are underlined. Circles represent unique bacterial genotypes. The color of the circle
931 fill represents the time point at which a genotype was firstly detected. *OBC3* and *gac* mutants
932 are highlighted in blue and purple, respectively. Right) Muller plots depicting the dynamics of
933 mutant alleles during the evolutionary experiment. The Muller plots show the estimated
934 frequencies, by their height, of 33 mutations in the respective population over 6 experimental
935 cycles. Descendant genotypes are shown emerging from inside their respective ancestors. The
936 frequency of each mutation in the respective population can also be found in Table 2. Muller
937 plots are prepared using *ggmuller* in R (101).

938 **Figure 4**. Evolved, root-competent mutant *P. protegens* CHA0 strains are characterized by
939 enhanced swimming and impaired swarming. The accumulative effects of parallel and
940 sequential mutations on swimming motility (top panel) and swarming motility (bottom panel)
941 are shown. For each experimental line, strains carrying mutations that can be connected to
942 bacterial motility, or that are the ancestor or descendants of such mutant strains, were studied
943 using typical swimming and swarming assays on Cook's Cytophaga (CC) medium (95). The
944 colony area as a measure for bacterial motility was determined using ImageJ. The combined
945 data from three independent experiments is shown, and datapoints from each experiment can
946 be discerned by their respective shape. The sample size, n, varies per genotype as related
947 strains are combined on each plate, and we report the minimum and maximum n per replicate
948 experiment. For swimming assays, circles represent replicate 1 ($2 \leq n \leq 6$) (excludes *osmY*$^{104G>A}$

949 and *mraZ*$^{211A>G}$), triangles replicate 2 (3≤*n*≤15), and squares replicate 3 (4≤*n*≤20). For swarming
950 assays, circles represent replicate 1 (1≤*n*≤3) (excludes *osmY*$^{104G>A}$ and *mraZ*$^{211A>G}$), triangles
951 replicate 2 (2≤*n*≤6), and squares replicate 3 (7≤*n*≤35). Obvious outliers were removed after
952 pooling. Significant differences in motility between a genotype and its respective progenitor
953 were determined by unpaired t-test analysis (swimming, 7≤*n*≤41; swarming, 7≤*n*≤44; *α=0.05,
954 **α=0.01, ***α=0.005, ****α=0.001, ns = non-significant) and the result is shown above each
955 comparison. The genealogy of the mutations is shown below each experimental line
956 highlighting both parallel (branching) and sequential mutations. Colours depict the number of
957 acquired mutations relative to the ancestor (white), and these are light-grey: 1, grey: 2, light-
958 purple:3 and purple: 4.

959 **Figure 5**. Genotype-phenotype association analysis of evolved bacterial populations. Isolates
960 without mutations or only with synonymous mutations were excluded for this association
961 analysis **a)** Heatmap representing the performance of 30 life-history traits of six ancestral and
962 63 evolved isolates. The "ward.D2" method was used for hierarchical clustering, based on the
963 dissimilarities of the set of objects being clustered. The four clusters depicted on top are model-
964 predicted. Each column illustrates one bacterial life-history trait, indicated on the bottom of the
965 figure. Each row represents an isolate, indicated by its sample ID number and mutational
966 events. **b-e)** Principal component analysis was applied to generate a general symbolic index for
967 each model predicted cluster. The proxy values of each cluster were normalized, higher values
968 represent better performance. To illustrate better the accumulative effects of each mutational
969 event, ANOVA was applied to reveal the phenotypic change between each genotype and their
970 identified nearest ancestor in each line (only applied for genotypes that have been detected

971 more than once; n>=2). Asterisks alongside the mutational events indicate significant
972 differences (*α=0.05, **α=0.01, ***α=0.001; n.s. = non-significant). Colour-filled circles
973 represent bacterial genotypes, with different colours representing independent evolving
974 populations and the size denotes the number of replicates (n). The dashed line represents the
975 average value of the ancestor.

Supplementary Figure Legends:

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

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

Figure S3. Representative photographs of swimming (A) and swarming (B) motility assays. Photographs illustrate the experimental layout of motility assays in which every plate is inoculated with the ancestral strain alongside various sequential mutations clustered per experimental line. Genotypes are indicated in each respective corner. All plates were incubated with the right side up at 21 °C for 18 h and then photographed.

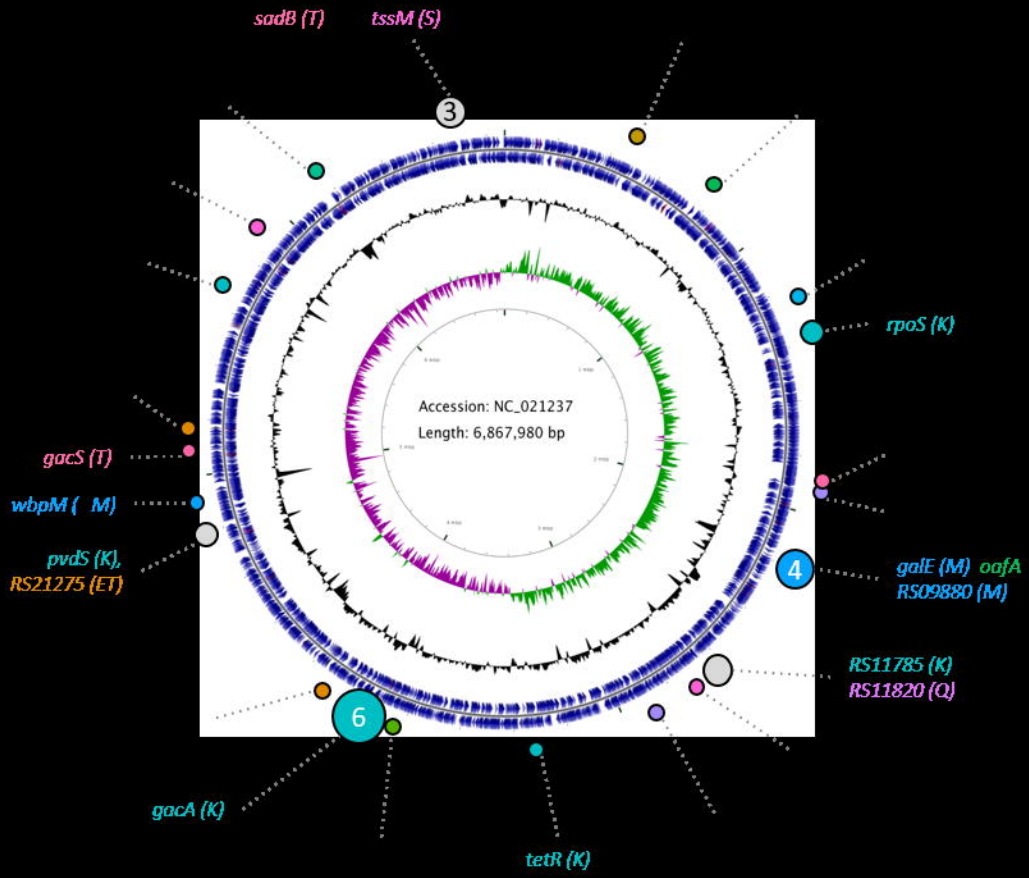
Figure S4. Selection of the best clustering model using Bayesian Information Criterion (BIC) using the *mclust* R package (97). A large BIC score indicates strong evidence for the corresponding model. The VEI model with 4 components best fits our data.

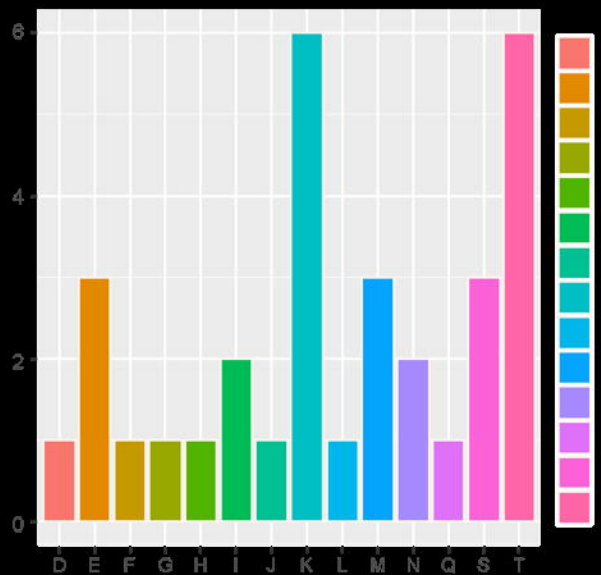
Figure S5. Coordinate axes transformation of each model predicted cluster using principal component analysis (PCA). **a.** PCA plot of social trait cluster (cluster 1); **b.** PCA plot of carbon utilization trait cluster (cluster 2); **c.** PCA plot of biocontrol trait cluster (cluster 3); **d.** PCA plot of stress resistance trait cluster (cluster 4). Each dot represents one isolate.

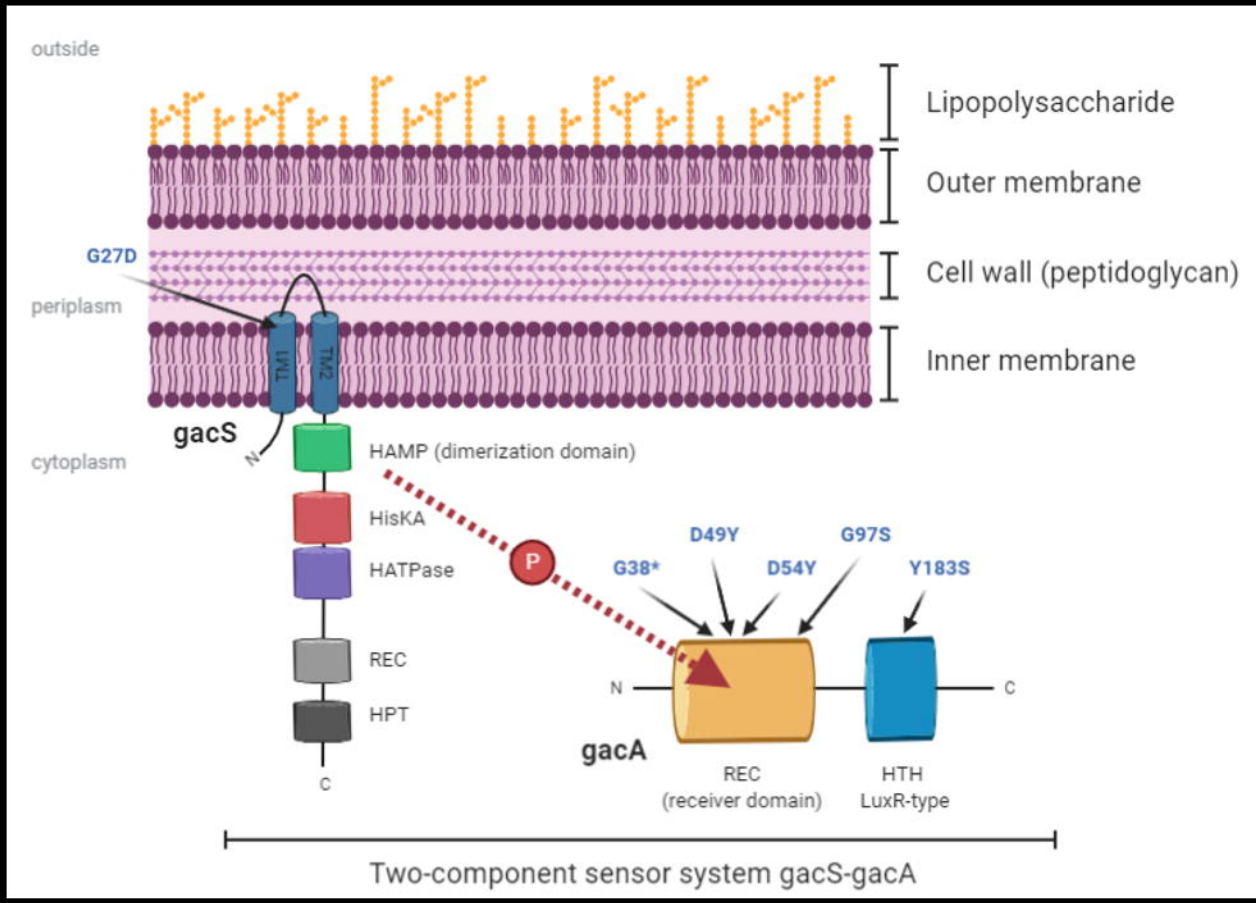
Supplementary Table Legends:

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

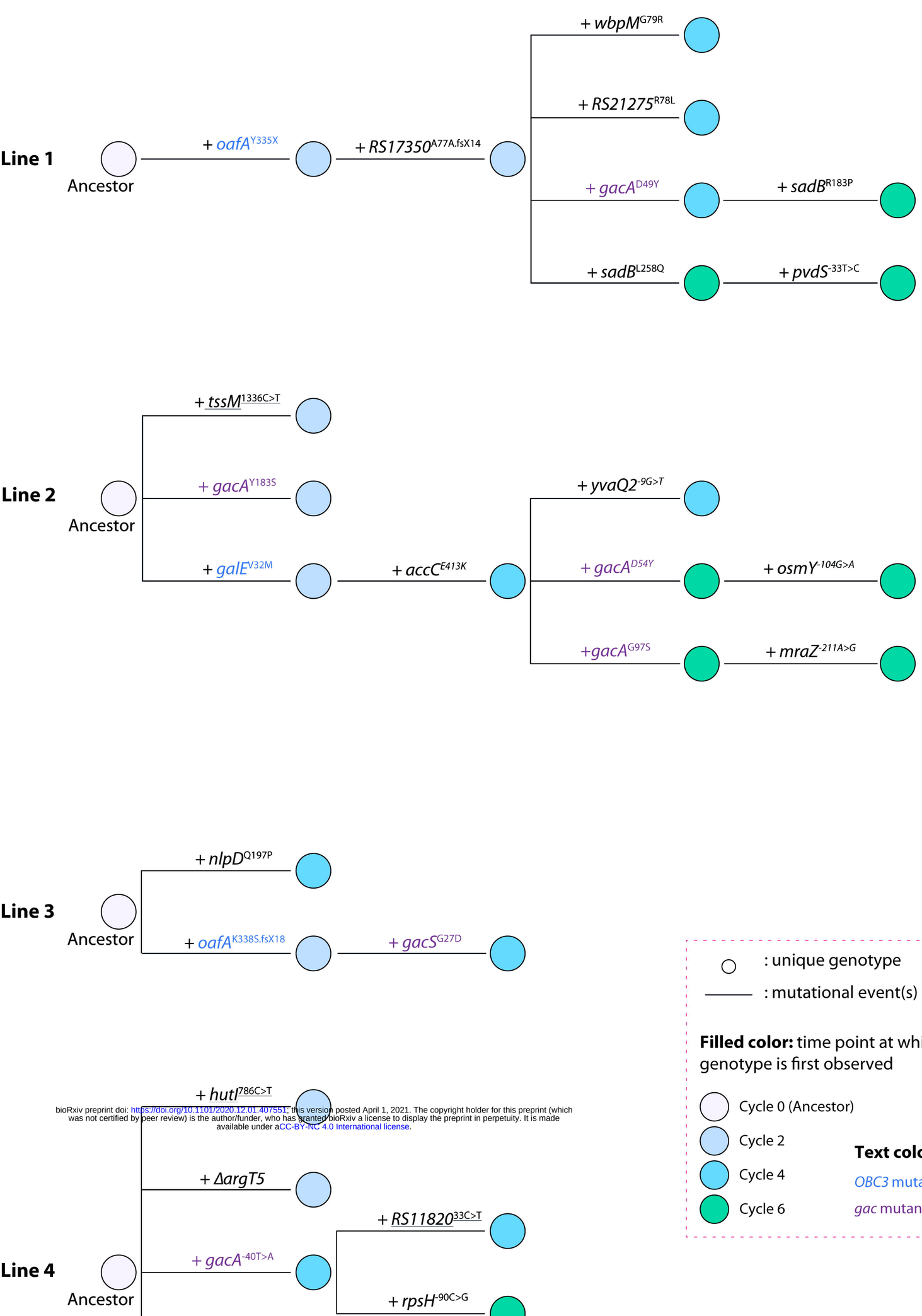
Table S2. Primers and probes used for high-resolution melting (HRM) analysis. For the two gacA mutants the same set of primers was used. Underlined bases indicate the position of the single nucleotide point (SNP) mutations within the probe sequences. ΔTm (℃) indicates the melting temperature difference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


sadB (T)
tssM (S)
3
rpoS (K)
Accession: NC_021237
Length: 6,867,980 bp
gacS (T)
wbpM (M)
pvdS (K),
RS21275 (ET)
4
galE (M) oafA
RS09880 (M)
RS11785 (K)
RS11820 (Q)
6
gacA (K)
tetR (K)


6
4
2
0
D E F G H I J K L M N Q S T


outside
Lipopolysaccharide
Outer membrane
G27D
periplasm
Cell wall (peptidoglycan)
TM1
TM2
Inner membrane
gacS
cytoplasm
N
HAMP (dimerization domain)
HisKA
HATPase
P
D49Y
G97S
G38*
D54Y
Y183S
REC
HPT
C
N
C
gacA
REC
(receiver domain)
HTH
LuxR-type
Two-component sensor system gacS-gacA


NC_021237.1
2,180,000
K338Sfs*18
2,190,000
Y82X
Y335X
V32M



: unique genotype

: mutational event(s)

Filled color: time point at which genotype is first observed

Cycle 0 (Ancestor)

Cycle 2

Cycle 4

Cycle 6

Text color:

OBC3 mutant

gac mutant