Evolutionary forecasting of phenotypic and genetic outcomes of experimental evolution in *Pseudomonas*

Peter A. Lind [1]

[1] Department of Molecular Biology, Umeå University, Umeå, Sweden.

Correspondence and requests for materials should be addressed to
Peter A. Lind, Dept. Molecular Biology, Umeå University, SE-901 87 Umeå, Sweden.

email: peter.lind@umu.se

Keywords: *Pseudomonas*, experimental evolution, genetic architecture, c-di-GMP, evolutionary predictability, evolutionary forecasting

Short title: Forecasting experimental evolution

Impact statement: Conservation of genotype-to-phenotype maps allows successful prediction of short-term evolution in *P. protegens* Pf-5 and lays the foundation for evolutionary forecasting in other *Pseudomonas.*

Abstract

Experimental evolution is often highly repeatable, but the underlying causes are generally unknown, which prevents extension of evolutionary forecasts to related species. Data on adaptive phenotypes, mutation rates and targets from the *Pseudomonas fluorescens* SBW25 Wrinkly Spreader system combined with mathematical models of the genotype-to-phenotype map allowed evolutionary forecasts to be made for several related *Pseudomonas* species. Predicted outcomes of experimental evolution in terms of phenotype, types of mutations, relative rates of pathways and mutational targets were then tested in *Pseudomonas protegens* Pf-5. As predicted, most mutations were found in three specific regulatory pathways resulting in increased production of Pel exopolysaccharide. Mutations were, as predicted, mainly found to disrupt negative regulation with a smaller number in upstream promoter regions. Mutated regions in proteins could also be predicted, but most mutations were not identical to those previously found. This study demonstrates the potential of short-term evolutionary forecasting in experimental populations.

Introduction

An increasing number of experimental evolution studies, primarily using microbes, have provided insights into many fundamental questions in evolutionary biology including the repeatability of evolutionary processes (Barrick and Lenski 2013; Jerison and Desai 2015; Long, et al. 2015; Orgogozo 2015; Blount, et al. 2018). Given the ability to control environmental conditions as well as population size and the use of a single asexual organism, such studies could provide an ideal test of our ability to predict evolutionary outcomes in simplified model systems. High repeatability on both phenotypic and genetic level have been observed in a large number of experimental evolution studies (Wichman, et al. 1999; Conrad, et al. 2009; Lee and Marx 2012; Tenaillon, et al. 2012; Barrick and Lenski 2013; Ferguson, et al. 2013; Herron and Doebeli 2013; Blank, et al. 2014; McElroy, et al. 2014; Fraebel, et al. 2017; Kram, et al. 2017; Blount, et al. 2018; Knöppel, et al. 2018), but it has become clear that high repeatability alone is not sufficient for testing evolutionary predictability beyond the prediction that under identical conditions the same evolutionary outcome is likely. The difficulties of moving from repeatability to predictability are largely a result of the lack of knowledge of the genotype-phenotype-fitness map (Figure 1).

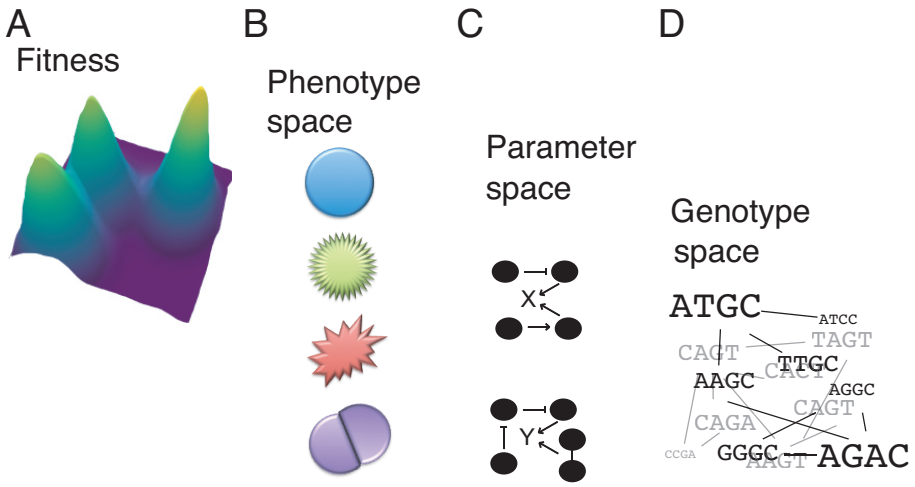


Figure 1. Prediction of the adaptive outcomes of experimental evolution requires understanding of how genotypes map to phenotypes and fitness. The differing capacity of genes to translate genotypic variation into phenotypic variation and differences in mutation rates can introduce biases in the production of phenotypic

variation. Natural selection, genetic architecture and mutational biases can both increase and decrease the predictability of evolution depending on if they can be recognized beforehand and included into evolutionary forecasting models. (**A**) **Fitness**. Mutants that increase to high frequencies in the population are all expected to have increased fitness as drift is negligible at population sizes typically used for adaptation experiments with microbes. Much effort has been put into characterizing the distribution of fitness effects of beneficial mutations, but the shapes of the distribution and magnitudes of the fittest mutations appear to be highly context dependent. Many different phenotypes are typically adaptive during experimental evolution, but in most cases they are not known beforehand and their relative fitness cannot be predicted. Relative fitness is, in most cases, also expected to be highly dependent on external environment including the frequency of other adaptive mutants, which means that even small changes to experimental protocols can lead to differences in outcomes. (**B**) **Phenotype space**. Each of the adaptive phenotypes can usually be realized by mutations in different positions and in different genes, but distinct phenotypes are expected to have similar fitness regardless of genetic foundations, which can simplify predictions. Depending on the genetic architecture underlying each trait, which is often unknown, adaptive phenotypes are produced at different rates. (**C**) **Parameter space**. The adaptive phenotypes are caused by changes in the molecular networks of cells, which are also influenced by the external environment. If the wiring of a molecular network underpinning an adaptive phenotype is well understood, parameterization of the system is possible and predictive models can be formulated. Mutations can cause functional effects on gene products, but mutations in some genes are more likely to lead to phenotypic variation. This can for example be due to differences in mutational robustness of the gene products themselves or their functions in regulatory networks. (**D**) **Genotype space**. Mutation rates are not uniform across the genome and mutational hot spots can lead to bias in the number of mutants producing relevant changes in parameter space and causing new phenotypes that are presented for natural selection to act upon. The current understanding of the distribution of mutation rates is limited and computational predictions have not yet been described. Adding information about well-characterized mutational hot spots, including indels in homonucleotide tracts or deletion and duplication between sequence homologies, could possibly improve prediction compared to a null model with uniform rates. Alternatively experimental

data might be incorporated into models of mutation rates to improve predictions (Lind, et al. 2019).

There are several problems that need to be solved to develop a model system for true testing of predictive ability. In some cases adaptive mutations are highly strain specific, so that for example adaptation of different strains to a specific environment will produce different results. This is sometime due to the long history of sub-culturing under laboratory conditions combined with rounds of mutagenesis that has caused, for example, many *E. coli* and *Salmonella* strains to accumulate diverse mutations, some of which are rapidly compensated by secondary mutations restoring fitness (Barrick, et al. 2009; Tenaillon, et al. 2012; Knöppel, et al. 2018). Thus, in many cases conclusions from one strain cannot be extended to another because of differences in their genotype-to-phenotype maps (Figure 1C).

Another problem for testing predictability is that in many cases it is not possible to design an experiment where one specific selective pressure is dominant. For example experiments with intended adaptation to high temperature (Tenaillon, et al. 2012) or freeze-thaw-growth cycles (Sleight, et al. 2008) result in similar mutations in *uspA*, which may indicate adaptation to the medium used (Knöppel, et al. 2018) or generally stressful conditions. Relatively minor changes in environmental conditions can also results in divergent mutational patterns (Deatherage, et al. 2017). This means that the range of possible adaptive phenotypes cannot be defined beforehand (Figure 1B) and that in many cases the phenotypes that solve the intended selective problem are outcompeted by other phenotypes with increased fitness (Figure 1A).

A highly specific selective pressure can be applied by selection for antitbiotic resistance and mutation targets are often highly conserved between different strains and species and between laboratory and natural populations (O'Neill, et al. 2006; Schenk, et al. 2012; Brandis, et al. 2015; Jahn, et al. 2017; Lukacisinova and Bollenbach 2017; Sommer, et al. 2017). However resistance phenotypes are typically explained solely by the molecular phenotype of a single protein and no alternative pathways to resistance are known resulting in a relatively simple parameter and genotype space (Figure 1C, Figure 1D). Thus the prediction will be identical for all species and it cannot provide a test of prediction from general principles. In many

133 cases mutants isolated after selection for high-level antibiotic resistance also lacks the
134 complexity that is inherent to many phenotypic traits where the genotype-to-
135 phenotype map involves a large number of functional interactions and complex
136 regulation (Figure 1C). This complexity comes to light in that mutations allowing
137 adaptation to new environments are commonly found in global regulators of gene
138 expression such as genes involved in the stringent response, DNA binding proteins,
139 supercoiling and core genes for RNA and protein synthesis (Barrick, et al. 2009;
140 Conrad, et al. 2009; Kishimoto, et al. 2010; Tenaillon, et al. 2012; Herron and
141 Doebeli 2013; Sandberg, et al. 2014; LaCroix, et al. 2015; Deatherage, et al. 2017).
142 The physiological effects of these mutations are diverse, sometimes affecting the
143 expression of hundreds of genes making the elucidation of the molecular
144 underpinnings of the adaptive phenotype (Figure 1C, 1D) extremely complex and thus
145 difficult to use for predictive modeling.
146
147 The wrinkly spreader model in *P. fluorescens* SBW25 (hereafter SBW25) is one of
148 the best-characterized experimental evolution systems and has several properties that
149 could make it possible to extend knowledge and principles from this species to related
150 species (Rainey and Travisano 1998; Spiers, et al. 2002; Spiers, et al. 2003; Spiers
151 and Rainey 2005; Goymer, et al. 2006; Bantinaki, et al. 2007; McDonald, et al. 2009;
152 Silby, et al. 2009; Ferguson, et al. 2013; Lind, et al. 2015, 2017b; Lind, et al. 2019).
153 When the wild type SBW25 is placed into a static growth tube the oxygen in the
154 medium is rapidly consumed by growing bacteria (Figure 2A). However oxygen
155 levels at the surface are high and mutants that are able to colonize the air-liquid
156 interface have a major growth advantage and rapidly increase in frequency (Figure
157 2A). Several phenotypic solutions to air-liquid interface colonization, all involving
158 increased cell-cell adhesion, have been described and are distinguishable by their
159 colony morphology on agar plates (Figure 2A, 2B) (Rainey and Travisano 1998;
160 Ferguson, et al. 2013; Lind, et al. 2017b). The most successful of these is the Wrinkly
161 Spreader (WS) (Ferguson, et al. 2013; Lind, et al. 2017b) that overproduces a
162 cellulosic polymer that is the main structural component of the mat at the air-liquid
163 interface (Spiers, et al. 2002; Spiers, et al. 2003). The WS phenotype is caused by
164 mutational activation of c-di-GMP production by a diguanylate cyclase (DGC)
165 (Figure 2C) (Goymer, et al. 2006). While many different DGCs can be activated to
166 reach the WS phenotype, some are greatly overrepresented due to larger mutational

target sizes leading to a hierarchy of genetic routes to WS (Figure 2D) (McDonald, et al. 2009; Lind, et al. 2015). The genotype-to-phenotype map to WS has been characterized in detail (Goymer, et al. 2006; McDonald, et al. 2009; Lind, et al. 2019) allowing the development of mathematical models of the three main pathways to WS (Wsp, Aws and Mws) and the prediction of evolutionary outcomes (Figure 2E) (Lind, et al. 2019).

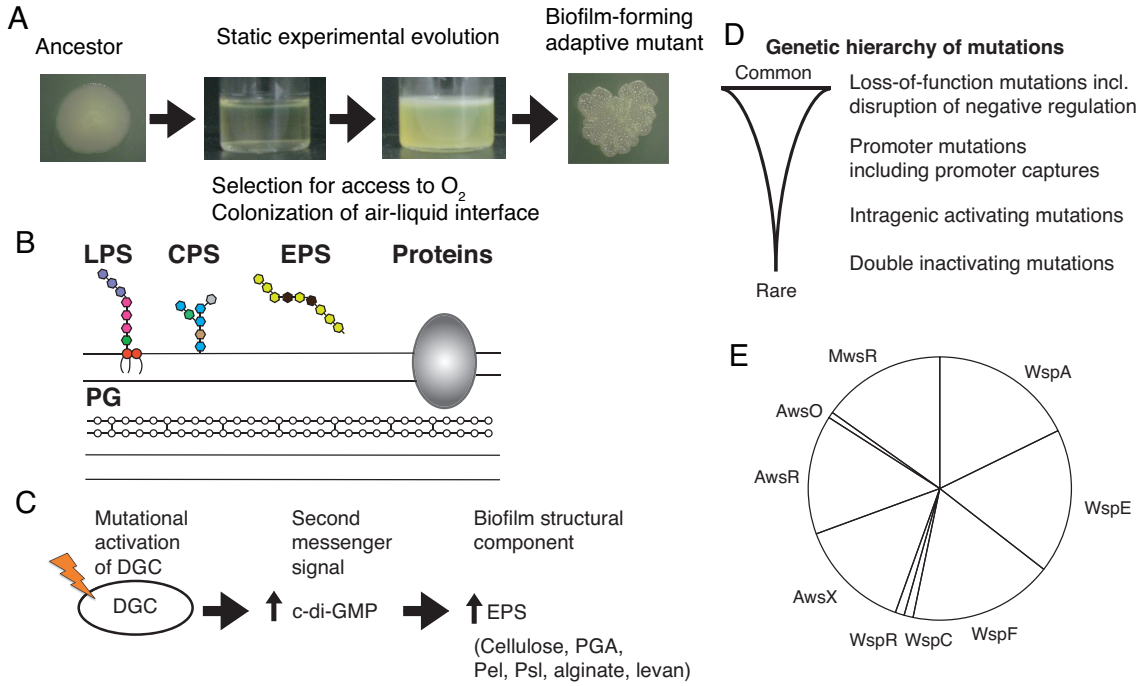


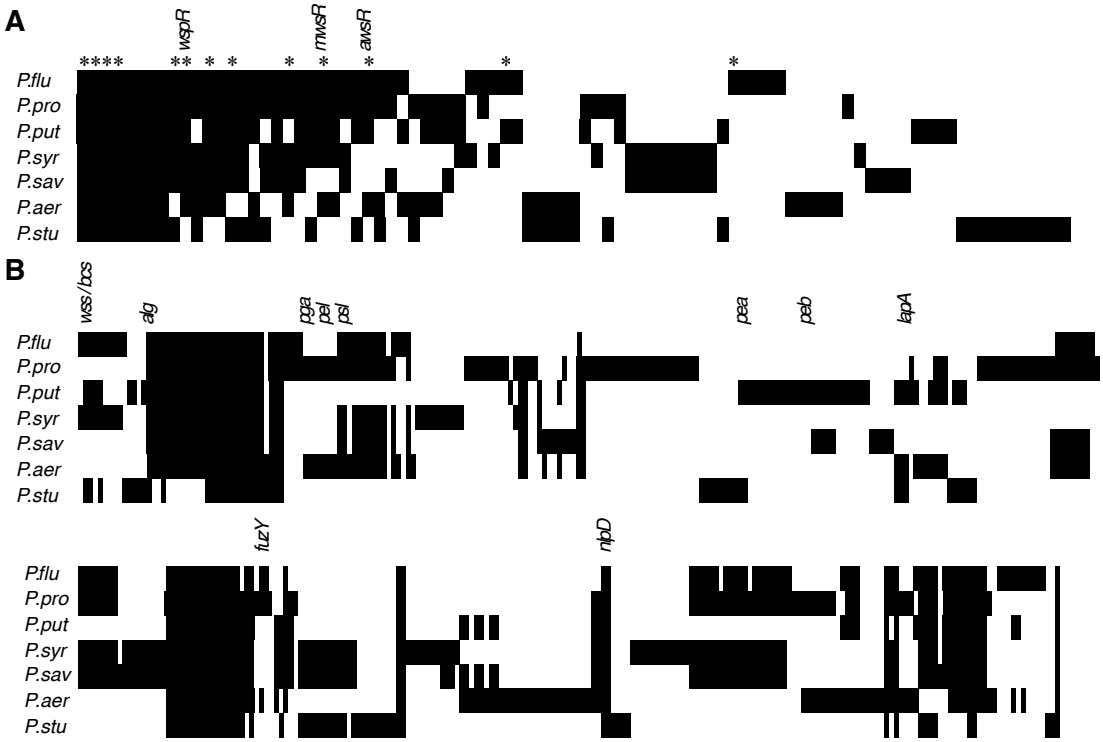
Figure 2. The *Pseudomonas fluorescens* SBW25 "wrinkly spreader" model system has several properties that could allow its extension to other species (**A**) The ancestral strain that has smooth colony morphology on agar plate is inoculated into static growth tubes and incubated for several days. Depletion of oxygen in the medium leads to competition for access to the oxygen-replete surface which is colonized by mutants with enhanced ability for cell-cell adherence and adherence to the wall of the tube. The most successful of these mutant types is the wrinkly spreader that has a distinctive colony morphology due to overproduction of exopolymeric substances (EPSs) of which a cellulosic polymer is the main structural component (Spiers, et al. 2002; Spiers, et al. 2003; Ferguson, et al. 2013; Lind, et al. 2017b) (**B**) The cell wall of Gram-negative bacteria, such as *Pseudomonas*, presents several components that may be used to increase cell-to-cell adhesion and attachment to surfaces. In addition to different EPSs, capsular polysaccharides (CPSs) and lipopolysaccharides (LPS) might also be used. A variety of different types of proteins including adhesins,

189 glycoproteins, fimbriae and pili are also potential phenotypic solutions and
190 incomplete cleavage of the peptidoglycan (PG) layer can lead to cell-chaining thereby
191 increasing cell-cell adhesion. In *Pseudomonas fluorescens* at least four alternative
192 distinct phenotypes are selected for at the surface, involving an alternative EPS, CPS,
193 LPS and PG, but they all have lower fitness than the WS type (Beaumont, et al. 2009;
194 Ferguson, et al. 2013; Gallie, et al. 2015; Lind, et al. 2017b). The wrinkly spreader
195 system is of intermediate complexity in that it allows isolation of a phenotypic subset
196 of adaptive mutants, those that are selected at the air-liquid interface and produce a
197 difference in colony morphology, rather than all mutants that increase fitness.
198 However there is also a large diversity of different genetic and phenotypic solutions to
199 the dominant adaptive challenge, which makes it of greater complexity than systems
200 based on single genes, which is often the case for experimental systems using strong
201 selection for antibiotic resistance. (**C**) The mutational causes of the high fitness WS
202 phenotype in wild type populations are found in either of three loci, *wspABCDEFR*,
203 *awsXRO* and *mwsR,* which all encode diguanylate cyclases that produce the second
204 messenger c-di-GMP (McDonald, et al. 2009) that is a conserved signal for EPS
205 production and biofilm formation in many bacterial species. In SBW25 the primary
206 EPS used is a cellulosic polymer. (**D**) While these three pathways account for >98%
207 of WS mutants in the wild type, 13 additional rare pathways were found when the
208 common ones are genetically deleted (Lind, et al. 2015). The large differences in
209 mutation rates to WS for the different pathways are explained mainly by their
210 different capacities to translate genotypic variation into phenotypic variation, i.e.
211 mutational target size, with the three common pathways being subject to negative
212 regulation, which when disrupted results in overproduction of c-di-GMP and the WS
213 phenotype (Lind, et al. 2015). The alternative phenotypic solutions are also caused by
214 inactivating mutations occurring at high rates (Ferguson, et al. 2013; Lind, et al.
215 2017b). The pathways to WS of intermediate frequency are activated by mutations to
216 promoter regions, including promoter captures and the most rare are those that require
217 specific activating mutations in the DGC or double mutations in two negative
218 regulators. The genetic underpinnings of several adaptive phenotypes have been
219 elucidated providing a mechanistic understanding of the effects of mutations and why
220 they are adaptive. (**E**) The three main pathways (Wsp, Aws, Mws) to WS are
221 particularly well-understood allowing formulation of mathematical models of the

222 molecular networks and prediction of the relative rates of use of the different
223 pathways and genes (Lind, et al. 2019).
224
225 This study makes initial forecasts of phenotypic and genetic evolutionary outcomes
226 after static experimental evolution for six *Pseudomonas* species based mainly on their
227 genome sequence and data from SBW25 (McDonald, et al. 2009; McDonald, et al.
228 2011; Ferguson, et al. 2013; Lind, et al. 2015, 2017b). Predictions of evolutionary
229 outcomes were then experimentally tested for the closely related species, *P. protegens*
230 Pf-5 (hereafter Pf-5) with a highly conserved genetic repertoire of DGCs but that
231 lacks the main structural component used by WS types in SBW25. Results show that
232 phenotypes, order of pathways used and types of mutations can be predicted and that
233 forecasts are robust to changes in environmental conditions.
234

235 **Results**

236 Six *Pseudomonas* species (Figure 3 legend) were chosen based on phylogenetic
237 diversity and their complement of DGCs and EPSs for a first round of predictions.
238 These species encode from none to all three of the main DGCs used in SBW25 and
239 only three species contain genes related to cellulose biosynthesis, the main EPS used
240 in SBW25 (Figure 3). Full details are available in Figure 3 - source data.



241

242 **Fig 3. Diversity of DGCs and biofilm-related genes for seven *Pseudomonas***
243 **species** (*P. fluorescens* SBW25, P. *protegens* Pf-5, *P. putida* KT2440, *P. syringae* pv.
244 tomato DC3000, *P. savastanoi* pv. phaseolicola 1448A, *P. aeruginosa* PAO1, *P.*
245 *stutzeri* ATCC 17588) **(A)** The seven species encode 251 putative DGCs, divided into
246 87 different homolog groups of which 8 are present in all genomes. WS mutations in
247 SBW25 have been found affecting 13 of these DGCs (marked with *) with an
248 additional nine that have been detected only in combinations with other mutations.
249 SBW25 and Pf-5 share 33 DGCs with 6 unique for each species. It should be noted
250 that not all DGCs are likely to be catalytically active. **(B)** Diversity of biofilm-related
251 genes including putative EPSs, LPS modification, cell chaining, adhesins and known
252 regulators. In SBW25 cellulose-based mats are most successful (encoded by the *wss*
253 operon). A secondary exopolysaccaride (PGA), encoded by *pgaABCD*, can also be
254 used to form a stable mat (Lind, et al. 2017b). Fuzzy spreaders (FS) forms rafts that
255 collapse after becoming too large and the mutational cause is inactivation of *fuzY*,
256 which results in a defect in lipopolysaccharide (LPS) modification (Ferguson, et al.
257 2013). Cell-chaining (CC) types have loss-of-function mutations in *nlpD* causing a
258 defect in cell division, which leads to segmented chains of cells that can form a weak
259 mat at the surface (Lind, et al. 2017b). The genomes of the *P. aeruginosa* strains PA7,
260 UCBPP-PA14 and LESB58 were also included in the analysis and results were in
261 most cases identical to PAO1 (not shown in Figure 3) with the exception of an
262 absence of homologues for EPS genes *pelA-D* and the DGCs PA2771 in PA14 and
263 PA3343 in PA7.
264
265 **Ecotype predictions**
266 Given the range of ways that cells can achieve increased adherence and surface
267 colonization by use of different EPSs, LPS modification and cell chaining as
268 demonstrated by the studies with SBW25 (Spiers, et al. 2002; Ferguson, et al. 2013;
269 Lind, et al. 2017b) all species are be expected to colonize the air-liquid interface if
270 access to oxygen is limiting for growth. This could be achieved simply by changes in
271 gene expression in the wild type, but for an experimental evolution study a mutational
272 solution is sought and environmental conditions are chosen so that the wild type strain
273 does not colonize the air-liquid interface. However changing environment presents a
274 further challenge because, as discussed above, it often leads to a different spectrum of
275 adaptive mutations. Thus a foundational requirement for an extended experimental

276 evolution system to be successful for different species is that the evolutionary
277 solutions are robust to differences in environmental conditions.
278
279 **Phenotype predictions**
280 Several different phenotypic solutions can be used to colonize the air-liquid interface
281 in SBW25 including at least two different EPSs (cellulose (WS) and PGA (PWS),
282 LPS modification (fuzzy spreaders FS) and cell chaining (CC) (Spiers, et al. 2002;
283 Ferguson, et al. 2013; Lind, et al. 2017b). However wrinkly spreaders that form
284 cellulose-based mats are superior and rapidly outcompete all other types (Ferguson, et
285 al. 2013; Lind, et al. 2017b). Based on the limited data available it is predicted that
286 cellulose-based biofilms are superior in other species as well and that they will be the
287 primary structural solution when available as for *P. syringae*, *P. putida* and *P.*
288 *stutzeri*. For the three species lacking genes for cellulose biosynthesis, other EPSs are
289 predicted to be used. Based on studies of *P. aeruginosa* the primary EPS required for
290 pellicle formation at the air-liquid interface in this species is Pel, encoded by the
291 *pelABCDEFG* operon, which is also present in the Pf-5 genome and is predicted to be
292 the primary phenotypic solution for these species. The genome of *P. savastanoi* lacks
293 genes for biosynthesis of cellulose and Pel as well as other EPSs that are known to be
294 able to support mat-formation, such as PGA and there is not sufficient data at this
295 point to make a prediction of which one is likely to be the primary phenotypic
296 solution.
297
298 Overexpression of EPSs used for mat-formation at the air-liquid interface is in
299 SBW25 and *P. aeruginosa* linked to mutations increasing c-di-GMP production rather
300 than mutations in the promoters of, or genes in, the EPS operons themselves. This can
301 be explained by the role of post-translation regulation by c-di-GMP in the production
302 of cellulose, Pel, PGA and alginate (Lee, et al. 2007; Römling, et al. 2013; Steiner, et
303 al. 2013; Morgan, et al. 2014; Liang 2015; Whitney, et al. 2015). In these cases
304 transcriptional up-regulation alone is not sufficient to cause overproduction because
305 of lack of a c-di-GMP signal. Possibly there is also an additional benefit to using
306 activation of the c-di-GMP network in that it reduces motility, which is not needed
307 when established at the air-liquid interface and which consumes large amount of
308 energy to sustain and thus is likely to be selected against (Koskiniemi, et al. 2012; Lee
309 and Marx 2012).

310

311 **Prediction of types of mutations**

312 Disabling mutations are expected to be more common than enabling mutations and
313 therefore the prediction is that most mutations will be in genes where loss-of-function
314 mutations produce an adaptive phenotype (Fig 2D) (Lind, et al. 2015). This is the case
315 for the large majority of mutations in SBW25 including those activating main DGCs
316 WspR, AwsR and MwsR, which are all under negative regulation, as well as
317 disruption of the genes underpinning the FS phenotype (*fuzY*, PFLU0478) and CC
318 phenotype (*nlpD*, PFLU1301) (McDonald, et al. 2009; Ferguson, et al. 2013; Lind, et
319 al. 2017b). Next in the hierarchy of mutations are promoter mutations, increasing
320 transcription, and promoter capture events (Lind, et al. 2015). Less common are
321 intragenic activating mutations that enable a particular function by for example
322 increase in catalytic activity or strengthening of interactions to another molecule or
323 another domain of the same protein (Lind, et al. 2015). Gene duplications occur at a
324 high rate and clearly have the ability to increase gene expression of DGCs, but they
325 have not yet been found to cause WS in SBW25, possibly because a two-fold increase
326 in gene expression is insufficient.

327

328 **Prediction of pathways used**

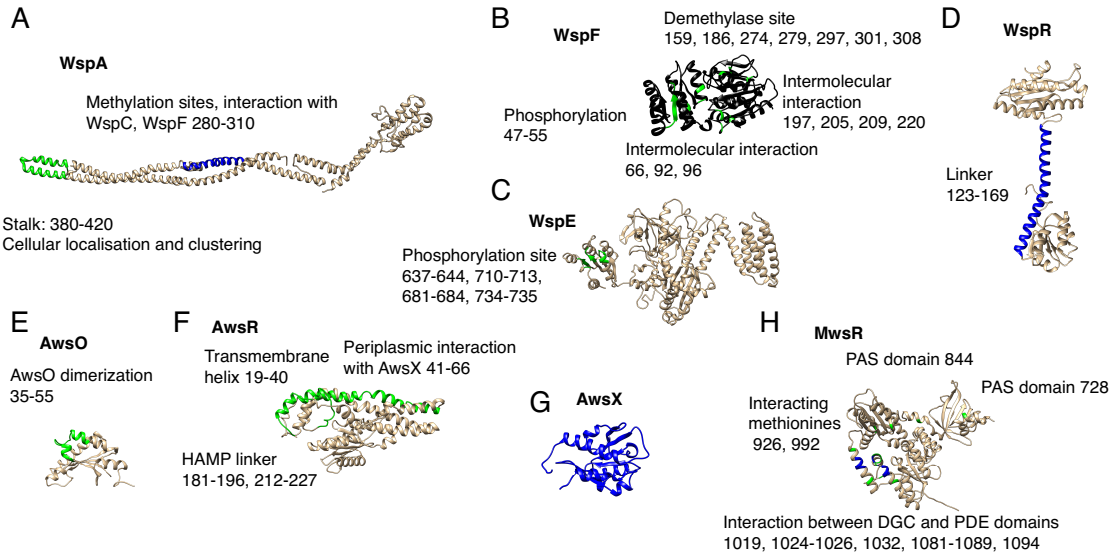
329 There are at least 16 different pathways to the WS phenotype in SBW25 with similar
330 fitness, but they are used at frequencies that vary over several orders of magnitude
331 based on the differing capacity to translate phenotypic variation into phenotypic
332 variation (Figure 2D) (Lind, et al. 2015). Mutations in three pathways, Wsp, Aws, and
333 Mws account for >98% of WS mutations and based on a detailed understanding of the
334 molecular functions of the genes involved of each pathways mathematical models
335 predicting at which relative rates the pathways should be used were constructed
336 (Figure 2E) (Lind, et al. 2019). The prediction results varies depending on the rates of
337 disabling and enabling mutations, but if it is assumed that disabling mutations are an
338 order of magnitude more common than enabling mutations the models predict that
339 Wsp will account for about 54%, Aws 30% and Mws 16% of the WS mutations
340 (Figure 2E) (Lind, et al. 2019).

341

342 When the three common pathways are deleted WS types evolve mainly by mutations
343 in PFLU0085, which contains an intragenic negative regulator region (Lind, et al.

344 2015), and this is expected to be the fourth most common pathways when present.
345 Less common promoter mutations will also appear at rates at least a magnitude lower,
346 but which DGCs that will be transcriptionally activated cannot be easily predicted
347 except for assuming it will be homologs of the ones used in SBW25. These DGCs
348 must be catalytically active and also be localized to the membrane (Farr, et al. 2017).
349 Possibly the subset of DGCs that are primarily activated by mutations to their
350 promoters is mainly determined by mutation rate and a higher mutation rate might be
351 caused by higher transcription and also influenced by gene direction (Sankar, et al.
352 2016). Most promoter capture deletion events are less than 5 kilobases (Lind, et al.
353 2015) in size and the lack of an alternative promoter relatively close upstream to the
354 DGC is likely to rule out these DGCs. The DGCs that can be activated by intragenic
355 activating mutations cannot now be predicted beyond the simple prediction that these
356 are the same genes as in SBW25 (Lind, et al. 2015).
357
358 **Prediction of mutated genes**
359 In addition to predicting the relative rates of the three main pathways, the previously
360 described mathematical model can also predict which proteins are likely to be
361 mutated (Lind, et al. 2019). High rates of WS mutations are predicted for WspF,
362 WspA, WspE, AwsX and AwsR and MwsR (Figure 2E). A significantly lower rate of
363 enabling mutations is also predicted to occur in WspC, WspR and AwsO (Figure 2E).
364 Despite the simplicity of null model it closely predicted the mutational targets in
365 SBW25 with equal rates for WspF, WspA and WspE and rare mutations in WspC and
366 WspR, suggesting that it is a useful null model also for other species (Lind, et al.
367 2019).
368
369 **Prediction of specific mutational targets and effects of mutations**
370 The level of parallelism at the nucleotide level between species is expected to be
371 dependent both on the number of possible mutations to WS and the degree of
372 functional conservation of the proteins involved that define the genotype-to-
373 phenotype map. Mutational hot spots are also expected to contribute to parallelism
374 when they are conserved, but reduce parallelism when they are not. Based on previous
375 analysis of patterns of mutations in SBW25 (McDonald, et al. 2009; McDonald, et al.
376 2011; Lind, et al. 2019) and homology modeling of protein structure using Phyre2

377 (Kelley, et al. 2015) regions expected to be mutated were predicted and the likely
378 molecular consequences of different mutations suggested (Figure 4).
379



380
381 **Figure 4.** Predicted mutational targets and proposed molecular effects. Black
382 represents any inactivating mutation including frame shifts, blue represents in frame
383 inactivating mutations, green represents amino acid substitutions. Numbers in
384 brackets refers to amino acid residue numbers in SBW25 (**A**) WspA – amino acid
385 substitutions are expected at the tip of the stalk and in-frame deletion of methylation
386 sites (**B**) WspF – any inactivating mutation is predicted, amino acid substitutions are
387 predicted only in areas where they disrupt intermolecular interactions (**C**) WspE –
388 amino acid substitutions are predicted near the phosphorylation site (**D**) WspR – small
389 in frame deletion and amino acid substitutions in the linker is predicted to cause
390 constitutive activation (**E**) AwsO – amino acid substitutions disrupting AwsO
391 dimerization is predicted to lead to increased binding to AwsX without the presence
392 of an activating signal (**F**) AwsR - amino acid substitutions in the periplasmic region
393 or transmembrane helix that disrupt the interaction with AwsX or to the HAMP linker
394 is predicted (**G**) AwsX – any inactivating mutation that keep the reading frame intact
395 and do not interfere with expression of downstream AwsR is predicted (**H**) MwsR –
396 mutations are predicted in the interface between the DGC and phosphodiesterase
397 domains and in the most C-terminal of the PAS domains resulting in constitutive
398 activation.
399
400 **Prediction of fitness effects of WS mutations**

401 While conservation of relative fitness of different phenotypic variants might be
402 expected it is less clear if the relative fitness of different DCG pathways and
403 mutations will be conserved between species. Despite this difficulty there might be a
404 way forward to predict the relative fitness of a large range of mutations with limited
405 experimental data. The distribution of fitness effects of new mutations have been
406 found to be bimodal for a large number of genes with different functions with one
407 mode close to neutrality and one corresponding to a complete loss of a particular
408 molecular function (Jacquier, et al. 2013; Jimenez, et al. 2013; Firnberg, et al. 2014;
409 Lind, et al. 2017a; Lundin, et al. 2017). Given that mutations that allow colonization
410 of the air-liquid interface have large phenotypic effects and are believed to also have
411 large effects on molecular function, often a complete disruption of an interaction,
412 adaptive mutations in the same region of a protein are likely to have similar fitness
413 effects. Thus, an approximation of the distribution of fitness effects could be possible
414 with relative few mutations for each gene. This is supported by the relatively small
415 number of WS mutants in SBW25 that have been characterized with sensitive fitness
416 assays and where mutations in the same gene typically have similar fitness effects
417 (Lind, et al. 2015; Lind, et al. 2019). If this assumption is true the distribution of
418 beneficial fitness effects is not continuous and the most advantageous mutations are
419 not predicted to be equally distributed between pathways or genes. Thus the
420 prediction would be that mutants isolated after experimental evolution were
421 concentrated to certain genes even if the mutational rate is similar so that although the
422 prediction from the null model is equal number of mutations for WspA, WspE and
423 WspF such distribution is unlikely to be found. While the mutation rates to WS for
424 the three genes are similar in SBW25, WspA mutants are rarely found after
425 experimental evolution due to their lower fitness (McDonald, et al. 2009; Lind, et al.
426 2019). There is however no *a priori* reason to expect that the relative fitness of
427 mutations in different genes or pathways will be conserved between species.
428
429 Inactivating mutations in *fuzY* and *nlpD* producing the alternative adaptive
430 phenotypes based on LPS modification or cell chaining were also found to have
431 similar fitness (Ferguson, et al. 2013; Farr 2015). Possibly there are other genes that
432 can be mutated with similar phenotypes, but that those mutants have lower fitness and
433 are outcompeted in SBW25. If relative fitness is not conserved between species this

434 could lead to high convergence on the phenotypic level but with completely different
435 genetic bases.
436
437 **Experimental test of forecasts in *Pseudomonas protegens* Pf-5**
438 In order to test the predictions presented above, *P. protegens* Pf-5 was used for a
439 parallel experimental evolution study under static conditions (Figure 1A). Its genome
440 encodes homologs of all except one of the DGCs used in SBW25 including all the
441 three common pathways to WS (Wsp, Aws, Mws). Thus the genetic predictions in
442 terms of types of mutations and mutated genes are nearly identical to SBW25 and the
443 mathematical null models can be directly applied. However Pf-5 lacks genes for
444 biosynthesis of cellulose meaning that if c-di-GMP overproduction is the main
445 pathway used, as predicted, an alternative EPS component must be utilized. The
446 experimental conditions were modified to test the robustness of predictions to changes
447 in media composition, temperature and cell wall material compared to those used in
448 the original SBW25 system (Materials and Methods). After experimental evolution
449 for five days, dilutions were spread on agar plates and then screened for colonies with
450 divergent colony morphology, characteristic of many phenotypes that colonize the air-
451 liquid interface.
452
453 **Genetic pathways**
454 In total 43 independent mutants were isolated and the causal mutations were
455 identified (Figure 5, Figure 5 – supporting data). As predicted by the null model the
456 majority (40/43) of mutations were associated with the Wsp, Aws, and Mws pathways
457 that are subject to negative regulation (Figure 1D). In addition the prediction that
458 promoter mutations would be the second most common type of mutation was
459 successful with two mutations found upstream of the *aws* operon, which were
460 predicted to disrupt the terminator of a high expression ribosomal RNA operon
461 representing a promoter capture event. Promoter mutations were also found upstream
462 of PFL_3078, which is the first gene of a putative EPS locus (PFL_3078-3093) that
463 has not previously been described and that is only present in closely related strains.
464 The operon encodes genes typical of exopolysaccharide biosynthetic operons making
465 it highly likely it encodes the main structural component used by these mutants.

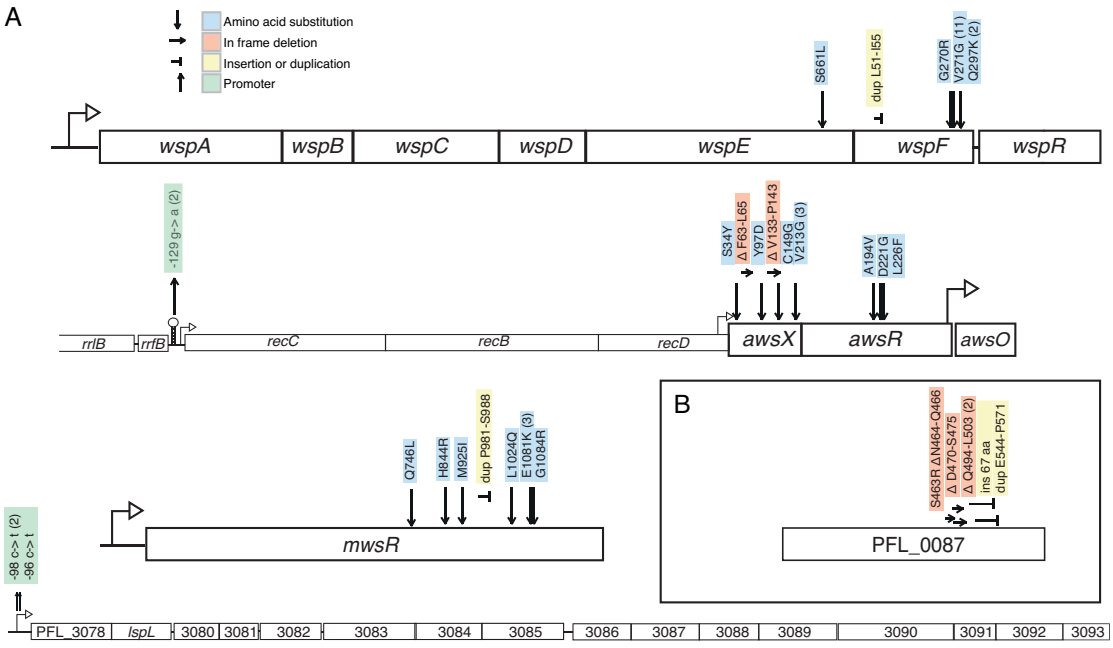


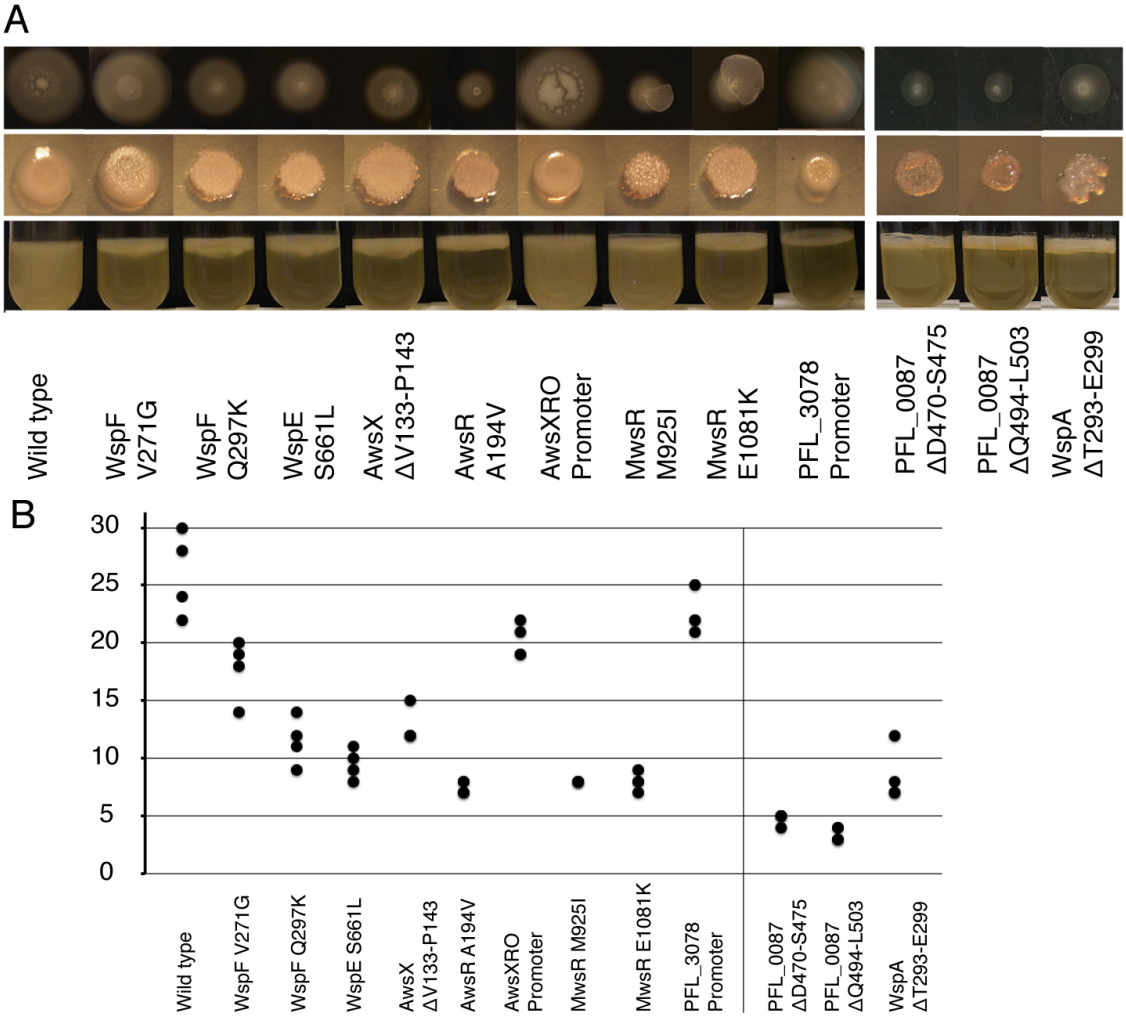
Figure 5. A. 43 independent mutants of wild type *Pseudomonas protegens* Pf-5 were isolated after experimental evolution based on their divergent colony morphology and mutations were identified in four operons. Numbers in brackets are the number of independent mutants found. Details are available in Figure 5 – source data. **B.** Experimental evolution with a Δ*wsp* Δ*aws* Δ*mws* triple deletion mutant resulted in WS types with mutations in the DGC PFL_0087.

The mathematical null model (Figure 2E) successfully predicted that of the three common pathways to WS, Wsp would be the most common one (16 mutants) followed by Aws (14) and then Mws (10). Mutations were predominately found in the negative regulators WspF (15 mutants) or AwsX (9), but also in interacting proteins WspE (1) and AwsR (3). Given that the mutational target size is estimated to be smaller for the interacting proteins (Figure 4) this is not surprising. No mutations were found in WspA despite a predicted high rate.

Mutations were predominantly found in predicted regions (Figure 4) for WspF, WspE, AwsX, AwsR and MwsR, but in most cases they were not identical to those in SBW25 (Figure 5 – supporting data). A mutational hot spot was apparent in WspF with 12 out of 15 mutations being identical V271G missense mutations. The previously described mutational hot spots in SBW25 in the *awsX* and *mwsR* genes (Lind, et al. 2019) appeared absent, demonstrating how mutation rate differences can

488 skew evolutionary outcomes even for closely related species with similar genetic
489 architecture.
490
491 To determine if there were also rare pathways to the WS phenotype the entire *wsp*,
492 *aws*, and *mws* operons were deleted and experimental evolution repeated as was
493 previously done for SBW25 (Lind, et al. 2015). Mutations in the DGC PFL_0087
494 accounted for six out of the seven WS types found (Figure 5B). This was also the
495 dominant pathway in the SBW25 Δ*wsp* Δ*aws* Δ*mws* strain where mutations in the
496 corresponding region of PFLU0085 were responsible for 47% of WS mutants. Thus
497 the fourth most common pathway is also the same for both species.
498
499 Mutations in WspA are predicted to be one of the major mutational routes to WS
500 based on the mathematical model (Lind, et al. 2019), but no mutations were found
501 either in this study or in SBW25 (McDonald, et al. 2009). However, when the
502 mutational spectrum of WS mutants was determined in the absence of selection for
503 growth at the air-liquid interface, *wspA* mutant occurred at rates similar to those of
504 WspE and WspF, as predicted by the model, and their low frequency after
505 experimental evolution could be explained by their lower fitness. To investigate if a
506 WspA mutation could cause WS in *P.protegens* a common deletion found in SBW25
507 (WspA T293-E299) was introduced and found to cause a WS phenotype and it is
508 included in the experiments described below.
509
510 **Phenotypic characterization**
511 In total 60 wells were inoculated for the wild type and subjected to experimental
512 evolution for five days after which air-liquid interface colonization was observed for
513 the majority of the wells. Mutants with clearly visible changes in colony morphology
514 were isolated from 43 wells. The experiment was repeated for the Δ*wsp* Δ*aws* Δ*mws*
515 triple deletion mutant and WS types were detected in 7 wells. Typically a single type
516 of divergent colonies was observed and one colony for each well was selected for
517 further characterization at random based on a pre-determined position on the agar
518 plate. Representative mutations were reconstructed using an allelic exchange protocol
519 to determine that the mutations are the sole cause of the air-liquid interface
520 colonization and colony phenotypes and to exclude the influence of secondary
521 mutations (Figure 6A) before further characterization.

522



523

524 **Figure 6.** Phenotypic characterization of reconstructed mutants. (**A**) Motility, colony
525 morphology and air-liquid interface colonization (**B**) Motility assay. As expected if
526 the c-di-GMP network is activated motility was reduced for most mutants with the
527 exception of PFL_3078 that is not expected to have increased c-di-GMP levels and
528 the AwsXRO promoter capture.

529

530 The lack of cellulose biosynthetic genes also shows that these ecotypes can evolve by
531 different phenotypes than in SBW25. Two clearly different phenotypes were observed
532 with one very similar to the original WS types in SBW25 with a clear motility defect
533 and mutations in the Wsp, Aws, Mws and PFL_0087 pathways (Figure 6A, 6B). The
534 other type was less wrinkly, had similar motility as the wild type and promoter
535 mutations upstream of the PFL_3078-3093 operon (Figure 6A, 6B).

536

537 **Fitness of adaptive mutants**

538 Two types of fitness assays were performed, similarly as previously described (Lind,
539 et al. 2015), to measure differences in fitness between the different WS mutants and
540 the alternative phenotypic solution with the mutation upstream of PFL_3078. The first
541 assay measures "invasion fitness" where the mutant is allowed to invade a wild type
542 population from an initial frequency of 1%. This confirms that the mutations are
543 adaptive and that mutants can colonize the air-liquid interface. The invasion assays
544 showed that all reconstructed mutants could rapidly invade an ancestral wild type
545 population (Figure 7A, Figure 7 – source data). Although there were significant
546 differences between selection coefficients of the mutants (one-way ANOVA $p <$
547 0.0001), no mutant was significantly different from the most common mutant (WspF
548 V271G, two-tailed *t*-test $p > 0.01$).
549
550 The second fitness assay measures "competition fitness" and here each mutant is
551 instead mixed 1:1 with the most common WS type (WspF V271G) at the start of the
552 competition. The competition assay showed that the ancestral wild type was rapidly
553 outcompeted by the mutants also at a 1:1 initial ratio (Figure 7B). There was
554 significant variation in fitness between the WS mutants (one-way ANOVA $p <$
555 0.0001) and the AwsX had significantly lower selection coefficient (two-tailed *t*-test p
556 $= 0.009$) compared to the reference WspF V271G and one of the MwsR mutants
557 (E1081K) had significantly higher selection coefficient (two-tailed *t*-test $p = 0.001$).
558 The alternative phenotypic solution used by the PFL_3078 promoter mutant resulted
559 in the lowest fitness ($s = -0.1$, two-tailed *t*-test $p < 0.0001$) meaning that it is expected
560 to be rapidly outcompeted by the WS mutants (Figure 7B). The PFL_0087 mutants
561 that were only found when the common pathways were deleted had lower fitness
562 (two-tailed *t*-tests $p = 0.0005$, $p = 0.001$) and this was also true for the *wspA* mutant
563 (two-tailed *t*-test $p = 0.002$), which could explain why these were not found in the
564 wild type population after experimental evolution.

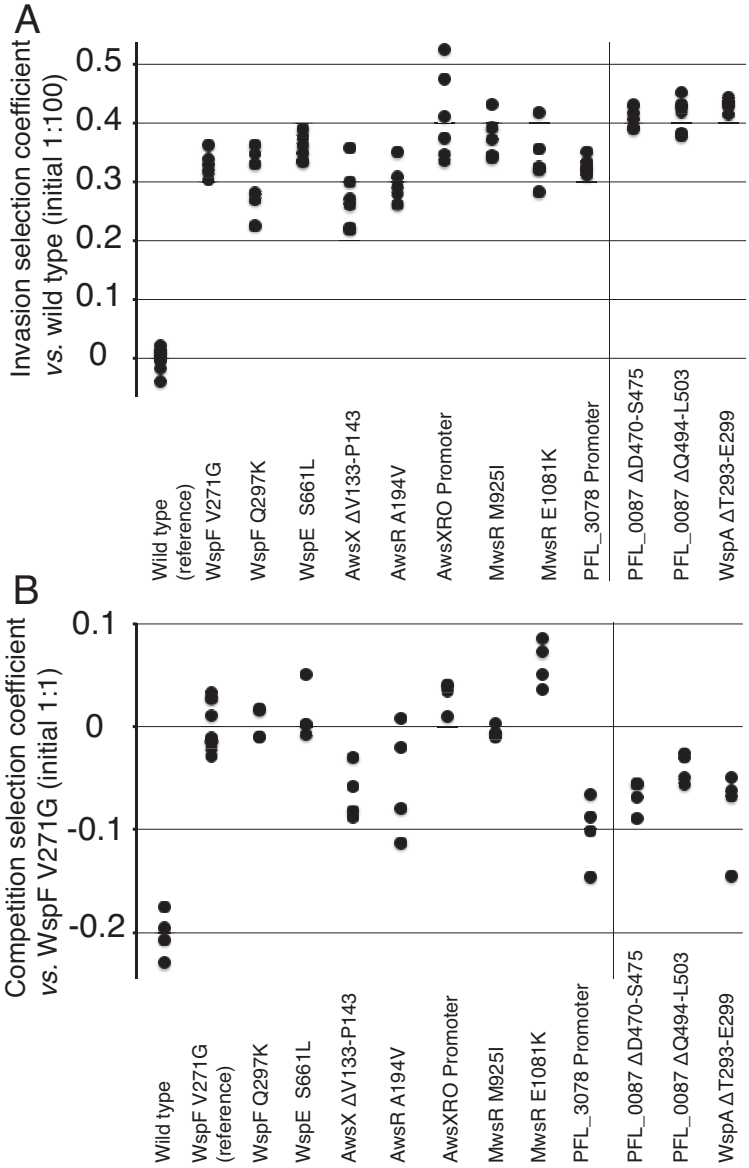


Figure 7. Fitness of reconstructed *P. protegens* Pf-5 WS mutants was measured in pairwise competitions. **(A)** Invasion fitness was measured relative a dominant ancestral wild type strain with a 1:100 initial ratio to show that mutations were adaptive and that they can increase from rare to colonize the air-liquid interface. Six independent competitions were performed for each pair. **(B)** Competition fitness was measured relative the most common WS mutant (WspF V271G) in a 1:1 initial ratio to compare the fitness of different WS mutants and the alternative phenotypic solution. Four independent competitions were performed for each pair.

Identification of EPS used for air-liquid colonization

SBW25 WS mutants use cellulose as the main structural component, but even though there is high parallelism at the genetic level for Pf-5 WS mutants this cannot be the

578 case at the phenotypic level as its genome does not encode genes for cellulose
579 biosynthesis. Given that production of Pel exopolysaccharide has been shown to be
580 induced by mutations in *wspF* in *P. aeruginosa* (Hickman, et al. 2005) and that Pel in
581 this species is required for pellicle formation under static growth (Friedman and
582 Kolter 2004) this was predicted to be the main structural component used by Pf-5. To
583 test this prediction the *pel* operon (PFL_2972-PFL_2978) was deleted from Pf-5 and
584 combined with previously characterized WS mutations and fitness was measured.
585 Both invasion fitness (Figure 8A) and competition fitness (Figure 8B) was
586 significantly lower (two-tailed t-tests $p < 0.01$) compared to isogenic strains with an
587 intact *pel* operon (Figure 7A, 7B) except invasion fitness for the AwsX mutant (two-
588 tailed t-tests $p = 0.08$, one outlier). This suggests that Pel polysaccharide serves as an
589 important structural component for colonizing the air-liquid interface and that its
590 production is activated by mutations leading to increased c-di-GMP levels. Although
591 deletion of *pel* in WS mutants resulted in less wrinkly colony morphology it did not
592 result in a smooth ancestral type. Neither did deletion of *pel* abolish the ability to
593 colonize the air liquid interface (Figure 8C) or the ability to invade wild type
594 populations (Figure 8A). This suggests that production of an additional EPS
595 component is induced by increased c-di-GMP levels caused by mutations in Wsp,
596 Aws, Mws and PFL_0087 at least in the absence of *pel*. When the cellulose
597 biosynthetic operon was deleted from SBW25, typical WS mutations in *wsp*, *aws*, and
598 *mws* resulted in air-liquid colonization by use of the alternative structural component
599 encoded by *pgaABCD* and subsequent deletion of the *pgaABCD* operon in these
600 mutants resulted in a wild type colony morphology (Lind, et al. 2017b). Deletion of
601 the *pgaABCD* operon (PFL_0161-PFL_0164) in Pf-5 strains with deletion of the *pel*
602 operon combined with WS mutations in either *wspF*, *awsX* and *mwsR* did not result in
603 a change in colony morphology or loss of ability to colonize the air-liquid interface,
604 which suggests that PGA is not the secondary structural component used or that yet
605 another EPS is also produced in response to increased c-di-GMP levels. As expected
606 if the motility defect observed for WS mutants are primarily caused by high c-di-
607 GMP levels rather than high production of Pel, the motility was also reduced for Wsp,
608 Aws and Mws mutants with the *pel* operon deleted (Figure 8D).

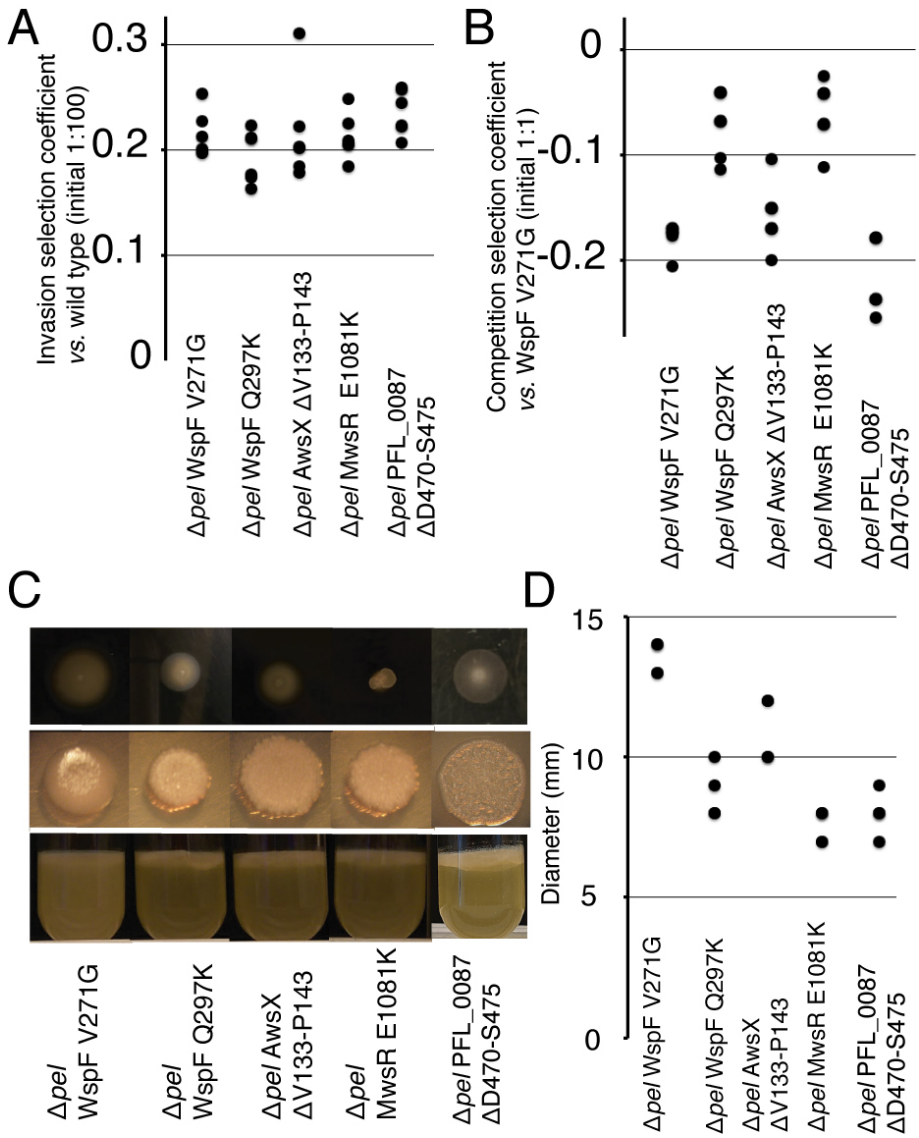


Fig 8. Contribution of *pel* to WS phenotype and fitness (**A**). Deletion of *pel* in WS mutants reduces invasion fitness. Fitness of reconstructed *P. protegens* Pf-5 WS mutants without the *pel* operon was measured in pairwise competitions. Invasion fitness was measured relative a dominant ancestral wild type strain with a 1:100 initial ratio. Six independent competitions were performed for each pair. **(B)** Deletion of *pel* in WS mutants reduces competition fitness. Competition fitness was measured relative the most common WS mutant (WspF V271G) in a 1:1 initial ratio. Four independent competitions were performed for each pair. (**C**) Deletion of *pel* in WS mutants did not result in ancestral smooth colony morphology or loss of ability to colonize the air-liquid interface suggesting a secondary EPS component is produced. (**D**). Deletion of *pel* did not restore motility showing that Pel overproduction is not the cause of the motility defect in WS mutants.

623 **Discussion**

624 The extension of the *P. fluorescens* SBW25 experimental evolution system to related
625 species shows promise for true testing of evolutionary forecasting method and
626 models. While there is a diversity of DGCs and EPSs between species leading to
627 differences in forecasts, the conserved role of c-di-GMP and limited number of
628 phenotypes allow the use of previous data to improve predictions and makes the
629 experimental system robust to changes in environmental conditions. The experimental
630 test of initial forecasts for *P. protegens* Pf-5 presented here provides support for the
631 ability to predict some aspects of both genetic and phenotypic evolution while
632 recognizing that the probability of specific mutations cannot in most cases be
633 predicted.
634
635 That experimental populations of *Pseudomonas* will colonize the air-liquid interface
636 when incubated under static condition is a prerequisite of extending the model. Given
637 that a range of phenotypic solutions is predicted to be available for all species the
638 evolution of such mutants for *P. protegens* is not surprising. The specific
639 environmental conditions used for experimental evolution often have a major impact
640 on evolutionary outcomes and is also likely to influence relative fitness and possibly
641 mutational biases also in the WS system. However despite major changes in growth
642 medium, temperature and material and physical dimensions of the growth vessel,
643 predictions on both the genetic and phenotypic levels proved successful
644 demonstrating robustness to environmental change and the establishment of a
645 dominant selective pressure, i.e. access to oxygen solved by air-liquid interface
646 colonization.
647
648 Phenotypic predictions of the structural basis supporting air-liquid colonization is
649 challenging given the limited previous experimental data. For SBW25 cellulose-based
650 solutions are superior in fitness, but for Pf-5 this solution in not available. The
651 prediction that overproduction of structural exopolysaccharides, rather than fuzzy,
652 cell-chaining or mucoid types, would be the primary solution was successful. One of
653 the two phenotypes found here used the Pel EPS, which could be predicted based on
654 its role in *P. aeruginosa*. However it appears to use a secondary EPS as well, that
655 remains to be identified, given that mutants lacking Pel but with activated DGCs still

656 colonize the air-liquid interface and have a distinct colony morphology. The second
657 phenotype used another EPS, encoded by PFL_3078-3093, which had not previously
658 be described and given that several EPS loci are usually encoded in *Pseudomonas*
659 genomes its use could not be predicted. However, repeating experimental evolution
660 using other *Pseudomonas* species is likely to provide more information about which
661 EPSs can be used to colonize the air-liquid interface and their relative fitness to allow
662 improved phenotypic predictions. Deletion of the *pel* operon, the unidentified
663 secondary EPS and PFL_3078-3093 and subsequent experimental evolution could
664 reveal less fit phenotypic solutions that are expected to exist including fuzzy types
665 caused by defects in LPS modification, cell-chaining types with defects in cell
666 division, adhesive proteins or mucoid types using alginate or levan, two EPSs with
667 lower structural stability.
668
669 The general prediction of types of mutations, as described in the hierarchy in Figure
670 2D, was also successful although the relatively few mutants identified here did not
671 allow for detection of rare activating mutations or double inactivating mutations. The
672 majority of mutations were loss-of-function mutations in negative regulators or
673 interacting proteins followed by less common promoter mutations and promoter
674 captures. In contrast to SBW25, where all promoter mutations resulted in up-
675 regulation of DGCs, the mutation upstream of PFL3078-3093 demonstrates the
676 possibility of direct transcriptional activation of EPS components that are not under
677 post-translational control of c-di-GMP. Two identical mutations were found over 9 kb
678 upstream of the *aws* operon, in between a ribosomal RNA operon and the *recCBD*
679 operon, which encodes key genes for recombination. The molecular effects of these
680 mutations have not been further investigated, but the resulting WS phenotype is
681 dependent of the presence of the *aws* operon, deletion of which reversed the
682 phenotype. This is consistent with an up-regulation of c-di-GMP by AwsR
683 presumably caused by increased transcription. The mutation is located in the
684 predicted terminator of the ribosomal RNA operon and increased transcriptional read-
685 through could put the *aws* operon under control of a very strong *rrn* promoter that is
686 most highly transcribed during exponential growth. This could explain the relatively
687 mild colony morphology phenotype as well as high motility of this WS mutant
688 (Figure 6).
689

690 A mathematical null model that incorporates information about the Wsp, Aws and
691 Mws molecular networks (Null model IV in (Lind, et al. 2019)) successfully predicted
692 that Wsp would be the most commonly used pathway followed by Aws, and Mws the
693 most rare. However, the number of mutants isolated here is rather small and the high
694 frequency of Wsp mutants seems mainly to be caused by a mutational hot spot in
695 *wspF*. Still the prediction that the three pathways together would contribute the large
696 majority of adaptive mutations (40 out of 43) is not trivial given that in SBW25 at
697 least 13 additional pathways are available to the high fitness WS phenotype (Lind, et
698 al. 2015). It is also worth noting that direct use of mutation rate data from SBW25
699 (Lind, et al. 2019) would result in poorer predictions than the mathematical null
700 model due to a strong mutational hot spot in *aws*X in that species. In addition the
701 fourth most common pathway to WS, PFL_0087, could be predicted based on data
702 from SBW25.
703
704 For the multi-protein pathways Wsp and Aws, the mathematical model (Lind, et al.
705 2019) predicted (Figure 2E) that mutations would primarily be found in WspA,
706 WspE, WspF, AwsX and AwsR. Mutations were detected in all these except in WspA
707 and the majority was found in the negative regulators WspF and AwsX. WspA
708 mutations were not found in the original study in SBW25 either (McDonald, et al.
709 2009), but this was shown to be due to lower fitness relative WspF and WspE mutants
710 rather than a lower mutation rate to WS (Lind, et al. 2019). This is also a likely
711 explanation for the absence of WspA mutants here as well (Figure 7B), but it is not
712 clear if this fitness difference would be conserved in other species or if sometimes
713 WspA mutants are more fit. Thus the null model prediction of equal rates for WspA,
714 WspF and WspE is not changed for future experimental tests. Direct comparison
715 between competitive fitness of mutants in SBW25 and Pf-5 is not possible as it was
716 measured under different experimental conditions, against different reference mutants
717 and in most cases the mutations are not identical. However it is interesting to note that
718 for both species high fitness WS types have mutations in the same proteins (WspF,
719 WspE, MwsR) and low fitness WS types also appear in the same proteins (WspA,
720 AwsX, AwsO, PFLU0085/PFL_0087).
721
722 The molecular effects of the mutations found here are unknown, but knowledge from
723 SBW25 and *P. aeruginosa* and their positions in protein structure allowed some

724 predictions to be made. Inactivating mutations in the negative regulator WspF were
725 predicted to be either indels or missense mutations in four specific regions. Mutations
726 were found in two of the predicted regions, one in the vicinity to the methylesterase
727 active site where mutations are predicted to cause large disruptions in protein
728 structure and the other one directly disrupting the phosphorylation active site in the
729 signal receiver domain. No mutations were found in the surface exposed regions
730 hypothesized to be involved in interactions with WspA and WspE, which could be
731 due to differences in function between SBW25 and Pf-5 or simply that they appear at
732 lower frequency and would be detected if additional mutations were isolated. The sole
733 mutation in WspE is, as predicted, located in the direct vicinity of the phosphorylation
734 active site. Mutations in AwsX were amino acid substitutions throughout the gene as
735 well as in frame deletions inactivating the gene as predicted. Mutations in AwsR and
736 MwsR were also found in predicted regions, but no mutations were found in the small
737 periplasmic region of AwsR, which is the most commonly targeted region in SBW25.
738 Known mutational hot spots in *awsX*, *awsR* and *mwsR* in SBW25 (Lind, et al. 2019)
739 were not conserved in Pf-5 resulting in divergent spectra of mutations, while mutated
740 regions and predicted functional effects remain conserved between the two species.
741 Little is known about the molecular function of the putative DGC encoded by
742 PFL_0087/PFLU0085, but it is clear that a multitude of amino acid substitutions,
743 deletions and insertions in a more than 40 amino acids long region can lead to WS
744 (Lind, et al. 2015). Thus it functions as a small intragenic negative regulator region
745 that might be involved in oligomerization and loss of this interaction results in
746 constitutive activation of c-di-GMP production.
747
748 The diversity of phenotypic solutions observed after experimental evolution is
749 dependent on fitness differences between the phenotypes, but also on the rate of
750 which phenotypes are introduced by mutations, which is dependent on the genetic
751 architecture underlying the trait as well as mutational biases. The Pf-5 strain has at
752 least three DGC pathways (Wsp, Aws and Mws) that are subject to negative
753 regulation leading to prediction of a high rate of WS mutants, which are then expected
754 to outcompete other phenotypic solutions. If instead only one of these pathways were
755 present, a larger diversity of phenotypes would be expected to be observed with
756 relative fitness becoming less important as the first mutant that gains a foothold at the

757 air-liquid interface will have a large advantage and priority effects, i.e. being first,
758 will increasingly determine which adaptive mutants are observed.
759
760 Given that the mutational target upstream of PFL_3078-3093 is likely be relatively
761 small and that these mutants are rapidly outcompeted by all WS types tested, their
762 relatively high frequency (3/43) is unexpected. Possibly this is due to a higher
763 mutation rate at these sites (Sankar, et al. 2016) or that population structure limits
764 direct competition between these different phenotypes and reduces the importance of
765 relative fitness. In SBW25 low fitness phenotypes that colonize the air-liquid
766 interface based on LPS modification or cell-chaining are observed prior to the rise of
767 WS to high frequencies (Lind, et al. 2017b) due to the presence of mutational hot
768 spots in these genes which make these mutants appear early during the growth phase
769 despite their relatively small mutational targets (Ferguson, et al. 2013; Farr 2015;
770 Lind, et al. 2017b).
771
772 In partially predicting evolutionary outcomes in *P. protegens*, this work lays the
773 foundation for future tests of evolutionary forecasting in related *Pseudomonas* species
774 by clearly stating predictions on several different levels from phenotype down to
775 which specific regions of proteins are likely to be mutated. Given what is already
776 known about the effects of (for now) unpredictable mutational biases and differences
777 in fitness between different WS types many of the forecasts will inevitably fail.
778 However hopefully they will fail in interesting ways thereby revealing erroneous
779 assumptions. The ability to remove common genetic and phenotypic pathways
780 provides a unique opportunity to also find those pathways that evolution does not
781 commonly use. This is necessary to determine why forecasts fail and update the
782 predictive models for another cycle of prediction, experimental evolution and mutant
783 characterization that make it possible to use this iterative model to define the
784 information necessary to predict short-term evolutionary processes.
785

786 ## Materials and methods

787

788 **Strains and media**

789 *Pseudomonas protegens* Pf-5 (previously known as *P. fluorescens* Pf-5) and
790 derivatives thereof were used for all experimental evolution and phenotypic
791 characterization. *E. coli* DH5α was used for cloning PCR fragments for genetic
792 engineering (Paulsen et al). *P. protegens* Pf-5 was grown in tryptic soy broth
793 (Tryptone 17g, Soytone 3g, Glucose 2.5g, NaCl 5g, K_2HPO_4 2.5g per liter)
794 supplemented with 10 mM MgSO4 and 0.2% glycerol (TSBGM) for experimental
795 evolution and fitness assays. Lysogeny broth (LB) was used during genetic
796 engineering and LB without NaCl and supplemented with 8% sucrose was used for
797 counter-selection of *sacB* marker. Solid media were 1.5% agar added to LB or TSB
798 supplemented with 10 mM MgSO4, 0.2% glycerol and 10 mg/l Congo red. Motility
799 assays were conducted in 0.3% agar TSB supplemented with 10 mM MgSO4, 0.2%
800 glycerol. Kanamycin was used at 50 mg/l for *E. coli* or 80 mg/l for *P. protegens* and
801 gentamicin at 10 mg/l for *E. coli* or 15 mg/L for *P. protegens*. Selection plates for
802 cloning contained 5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside (X-gal) at 40
803 mg/l. 100 mg/L nitrofurantoin was used to inhibit growth of *E. coli* donor cells after
804 conjugation. All strains were stored at -80°C in LB with 10% DMSO.
805
806 **Experimental evolution**
807 30 central wells of a deep well plate (polypropylene, 1.1 mL, round walls, Axygen
808 Corning Life Sciences) were inoculated with approximately 10^3 cells each and
809 incubated at 36°C for 5 days without shaking on two different occasions. Suitable
810 dilutions were plated on TSBGM plates with Congo red after 5 days and incubated at
811 36°C for 48 h. Plates were screened for colonies with a visible difference in colony
812 morphology and one divergent colony per well were randomly selected based only on
813 its position on the agar plate. In total 43 independent mutants were streaked for single
814 cells twice before overnight growth in LB and freezing. An identical protocol was
815 used for the Δ*wsp* Δ*aws* Δ*mws* strain.
816
817 **Genome sequencing**
818 Seven mutant strains that did not contain mutations in the *wspF* and *awsX* genes were
819 analyzed by genome resequencing. The strains had mutations in *awsR*, *mwsR*, *wspE,*
820 upstream PFL_3078 (2 strains) and in the intergenic region between *rrfB* and *recC*
821 upstream of the awsXRO operon. Genomic DNA was isolated with Genomic DNA

822 Purification Kit (Thermo Fisher). Sequencing libraries were prepared from 1µg DNA
823 using the TruSeq PCRfree DNA sample preparation kit (cat# FC- 121-3001/3002,
824 Illumina Inc.) targeting an insert size of 350bp. The library preparation was
825 performed according to the manufacturers' instructions (guide#15036187).
826 Sequencing was performed with MiSeq (Illumina Inc.) paired-end 300bp read length
827 and v3 sequencing chemistry.
828
829 Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The
830 facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for
831 Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research
832 Council and the Knut and Alice Wallenberg Foundation. Sequencing data were
833 analyzed with using Geneious v. 10.2.3 with reads assembled against the *P. protegens*
834 Pf-5 genome sequence (CP000076.1).
835
836 **Sanger sequencing**
837 Sanger sequencing were performed by GATC biotech and used to sequence candidate
838 genes to find adaptive mutations and to confirm reconstructed mutations. Primer
839 sequences are available in Table S1.
840
841 **Reconstruction of mutations**
842 Thirteen mutations representing all candidate genes found using Sanger or Illumina
843 sequencing as well as PFL_0087 and WspA mutations were reconstructed in the wild
844 type ancestral *P. protegens* Pf-5 to show that they are the cause of the adaptive
845 phenotype and to be able to assay their fitness effects without the risk of secondary
846 mutations that might have occurred during experimental evolution. A two-step allelic
847 replacement protocol was using to transfer the mutation into the ancestor. First a 1-2
848 kb fragment surrounding the putative adaptive mutations were amplified using PCR
849 (Phusion High- Fidelity DNA polymerase, Thermo Scientific) and ligated into the
850 multiple cloning site of the mobilizable pK18mobsac suicide plasmid (FJ437239)
851 using standard molecular techniques. The ligation mix was then transformed into
852 competent *E. coli* DH5α using heat shock. After confirmation of correct insert size by
853 PCR the plasmid was transferred to *P. protegens* Pf-5 by conjugation with the donor
854 strain and an *E. coli* strain carrying the conjugation helper plasmid pRK2013.
855 Cultures were grown overnight of the recipient *P. protegens* Pf-5 (20 ml per

856 conjugation at 30°C in LB) and 2 ml each of the donor and helper *E. coli* strains per
857 conjugation at 37°C in LB with kanamycin. The culture of *P. protegens* Pf-5 was heat
858 shocked for 10 minutes at 42°C prior to centrifugation at 4000 rpm for 10 minutes
859 and resuspension in a small volume of LB. Donor and helper cells were collected by
860 centrifugation 4000 rpm for 10 minutes, resuspended in LB, and mixed with the
861 concentrated recipient cells. After another round of centrifugation the conjugation mix
862 was resuspended in 50 µl LB and spread onto several spots on a LA plate followed by
863 incubation overnight at 30°C. Each spot of the conjugation mix was scraped of the
864 plate and resuspended in 200 µl LB each and plated on LA plates with kanamycin, to
865 select for transfer of the plasmid, and nitrofurantoin that prevents growth of the *E.*
866 *coli* donor and helper cells. The pK18mobsac plasmid has a pBR322 type origin and
867 cannot replicate in *P. protegens* Pf-5 and only cells where the plasmid has integrated
868 into the chromosome by homologous recombination, with the homology provided by
869 the cloned fragment, can grow in the presence of kanamycin. After streaking for
870 single cells on LA plates with kanamycin, the *P. protegens* Pf-5 strains with
871 integrated plasmids were grown overnight in LB at 30°C without antibiotics to allow
872 for double crossover homologous recombination resulting in loss of the integrated
873 plasmid. The plasmid also contains the *sacB* marker conferring sucrose sensitivity,
874 which allows for counter-selection by plating on LA plates with sucrose. Sucrose
875 resistant colonies were checked for loss of the kanamycin marker and DNA
876 sequencing of the cloned region to find strains with the reconstructed mutation and no
877 other mutations.
878 Deletion of the *wsp*, *aws*, *mws* and *pelABCDEFG* (PFL_2972-PFL_2978) regions was
879 accomplished using the same two-step allelic exchange protocol using SOE-PCR to
880 generated at fragment surrounding the operon as previously described (Ferguson, et
881 al. 2013; Farr 2015; Lind, et al. 2017b). Gene synthesis (Thermo Fisher) was used to
882 make DNA fragments used for deletion of PFL_0161-PFL_0164 and WspA T293-
883 E299. Primer sequences are available in Table S1.
884
885 **Fitness assays**
886 Two types of competition fitness assays were performed similarly to previously
887 described (Lind, et al. 2015). The first assay measures invasion fitness, where a
888 mutant is mixed 1:100 with the wild type ancestor, simulating early stages of air-

889 liquid interface colonization where a rare mutant establishes and grows at the surface
890 with no competition from other mutants. The second assay instead measures
891 competition fitness in a 1:1 competition against a reference mutant strain, which here
892 was chosen to be the WspF V271G mutant because it was the most commonly found
893 in the experimental evolution study and thus is highly successful either because of a
894 high rate of emergence, i.e. a mutational hot spot, or higher fitness than most other
895 WS mutants. In addition, the WspF V271G mutant has a temperature sensitive colony
896 morphology phenotype in that it is highly wrinkly at 30°C, but only have a very mild
897 phenotype when grown at room temperature, thus allowing it to be distinguishable
898 from both the smooth ancestor and all other wrinkly mutants isolated here.
899
900 Fluorescent reference strains of the wild type ancestor and the WspF V271G mutants
901 were created using a miniTn7 transposon (miniTn7(Gm) PA1/04/03 Gfp.AAV-a)
902 (Lambertsen, et al. 2004) that allows integration at a defined locus (attTn7) in the
903 chromosome. This allows the colonies to be distinguished not only by morphology
904 but also by fluorescence under blue/UV light and gentamicin resistance, which
905 provides a way to ascertain that secondary adaptive mutants that might occur during
906 the competition experiment do not bias the results (for example the ancestor could
907 evolve WS types or a WS mutant can evolve to cheat on the other type by inactivation
908 of EPS production or reduced c-di-GMP signalling). Introduction of the transposon
909 into *P. protegens* Pf-5 was performed by tri-parental conjugation from *E. coli* with
910 helper plasmids pRK2013 (conjugation helper) and pUX-BF13 containing the
911 transposase genes) using the same conjugation protocol described above.
912
913 The invasion assay was performed by mixing shaken overnight cultures of the
914 competitor 1:100 with the GFP-labeled reference ancestor followed by 1000-fold
915 dilution and static incubation at 36°C for 48 h in TSBGM medium in deep well plates
916 (1 ml per well, using only the central 60 wells). For the competition assay, the GFP-
917 labeled reference strain WspF 271G was mixed 1:1 with the competitor and diluted 6-
918 fold and grown for 4 h (shaken at 30°C), before plating to determine initial ratios, to
919 ensure the cells were in a similar physiological state at the start of the competition.
920 The competition cultures were then diluted 1000-fold in TSBGM medium and grown
921 in deep well plates (1 ml per well, using only the central 60 wells) static for 24 h at

922 36°C. Selection coefficients (s) were calculated as previously described (Dykhuizen
923 1990), where s = 0 is equal fitness, positive is increased fitness and negative is
924 decreased fitness relative to the reference strain. Briefly s is calculated as the change
925 in logarithmic ratio over time according to s = [ln(R(t)/R(0))]/[t], where R is the ratio
926 of mutant to reference and t is the number of generations of the entire population
927 during the experiment (estimated from viable counts). The cost of the fluorescent
928 marker were calculated from control competitions where the GFP-labeled reference
929 strains (wild type and WspF V271G) were competed against isogenic strains without
930 the marker and included in each plate under identical conditions during the fitness
931 assays and used to adjust the selection coefficients to compensate for the cost.
932
933 **Motility assays**
934 Swimming motility assays were performed in TSBGM plates with 0.3% agar (BD)
935 and the diameter was measured after 24 h of growth at room temperature. Each strain
936 was assayed in duplicates on two different plates.
937
938 **Bioinformatics analysis of DGCs and EPS genes**
939 Homologs for all DGCs in *P. fluorescens* SBW25 were found using the *Pseudomonas*
940 Ortholog Database at Pseudomonas.com (Winsor, et al. 2016). Blast-p searches for
941 GGDEF domains were performed to find remaining DGCs in the six *Pseudomonas*
942 species and their homologs again found using the *Pseudomonas* Ortholog Database
943 (Whiteside, et al. 2013) and manually inspected. Annotations (Pseudomonas.com. DB
944 version 17.2) were also searched for diguanylate cyclase and GGDEF. Not all DCCs
945 found are likely to have diguanylate cyclase activity, but given the difficulties of
946 predicting which of the partly degenerate active sites are likely to be inactive
947 combined with the possibilities of mutational activation during experimental
948 evolution, none were excluded.
949
950 There is no simple way to find all genes that can function as structural or regulatory
951 genes to allow colonization of the air-liquid interface. Thus the selection in Figure 3B
952 and Figure 3 – source data should not be considered complete. Putative EPS genes
953 were found using blastp searches with sequences from known exopolysaccharide
954 biosynthesis proteins including cellulose, PGA, Pel, Psl, Pea, Peb, alginate and levan.
955 Homologs were then found using the *Pseudomonas* Ortholog Database (Whiteside, et

956 al. 2013) at Pseudomonas.com (Winsor, et al. 2016). Annotations (Pseudomonas.com.
957 DB version 17.2) were also searched for glycosyltransferase, glycosyl transferase,
958 flippase, polysaccharide, lipopolysaccharide, polymerase, biofilm, adhesion and
959 adhesion. Based on previous work in SBW25 and literature searches a few additional
960 genes were added.
961

962 **Acknowledgements**

963 This work was supported by grants from Carl Tryggers Foundation for Scientific
964 Research and Magnus Bergvalls Foundation.
965

966 **Competing interests**

967 The author declares no competing interests.

968 **References**
969

970 Bantinaki E, Kassen R, Knight CG, Robinson Z, Spiers AJ, Rainey PB. 2007.
971 Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III.
972 Mutational origins of wrinkly spreader diversity. Genetics 176:441-453.
973 http://dx.doi.org/10.1534/genetics.106.069906
974 Barrick JE, Lenski RE. 2013. Genome dynamics during experimental evolution.
975 Nature reviews. Genetics 14:827-839. http://dx.doi.org/10.1038/nrg3564
976 Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF.
977 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia*
978 *coli*. Nature 461:1243-1247. http://dx.doi.org/10.1038/nature08480
979 Beaumont HJ, Gallie J, Kost C, Ferguson GC, Rainey PB. 2009. Experimental
980 evolution of bet hedging. Nature 462:90-93. http://dx.doi.org/10.1038/nature08504
981 Blank D, Wolf L, Ackermann M, Silander OK. 2014. The predictability of molecular
982 evolution during functional innovation. Proceedings of the National Academy of

983 Sciences of the United States of America 111:3044-3049.
984 http://dx.doi.org/10.1073/pnas.1318797111
985 Blount ZD, Lenski RE, Losos JB. 2018. Contingency and determinism in evolution:
986 Replaying life's tape. Science 362. http://dx.doi.org/10.1126/science.aam5979
987 Brandis G, Pietsch F, Alemayehu R, Hughes D. 2015. Comprehensive phenotypic
988 characterization of rifampicin resistance mutations in *Salmonella* provides insight into
989 the evolution of resistance in *Mycobacterium tuberculosis*. J Antimicrob Chemother
990 70:680-685. http://dx.doi.org/10.1093/jac/dku434
991 Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, Palsson BO. 2009.
992 Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-
993 term laboratory evolution in lactate minimal media reveals flexible selection of
994 adaptive mutations. Genome biology 10:R118. http://dx.doi.org/10.1186/gb-2009-10-
995 10-r118
996 Deatherage DE, Kepner JL, Bennett AF, Lenski RE, Barrick JE. 2017. Specificity of
997 genome evolution in experimental populations of Escherichia coli evolved at different
998 temperatures. Proceedings of the National Academy of Sciences of the United States
999 of America 114:E1904-E1912. http://dx.doi.org/10.1073/pnas.1616132114
1000 Dykhuizen DE. 1990. Experimental studies of natural selection in bacteria. Annu.
1001 Rev. Ecol. Syst. 21:373-398.
1002 Farr AD. 2015. Formalist features determining the tempo and mode of evolution in
1003 *Pseudomonas fluorescens* SBW25. [[Auckland, New Zealand]: Massey University.
1004 Farr AD, Remigi P, Rainey PB. 2017. Adaptive evolution by spontaneous domain
1005 fusion and protein relocalization. Nat Ecol Evol 1:1562-1568.
1006 http://dx.doi.org/10.1038/s41559-017-0283-7

1007 Ferguson GC, Bertels F, Rainey PB. 2013. Adaptive Divergence in Experimental
1008 Populations of *Pseudomonas fluorescens.* V. Insight into the Niche Specialist "Fuzzy
1009 Spreader" Compels Revision of the Model Pseudomonas Radiation. Genetics.
1010 http://dx.doi.org/10.1534/genetics.113.154948
1011 Firnberg E, Labonte JW, Gray JJ, Ostermeier M. 2014. A comprehensive, high-
1012 resolution map of a gene's fitness landscape. Molecular biology and evolution
1013 31:1581-1592. http://dx.doi.org/10.1093/molbev/msu081
1014 Fraebel DT, Mickalide H, Schnitkey D, Merritt J, Kuhlman TE, Kuehn S. 2017.
1015 Environment determines evolutionary trajectory in a constrained phenotypic space.
1016 Elife 6. http://dx.doi.org/10.7554/eLife.24669
1017 Friedman L, Kolter R. 2004. Genes involved in matrix formation in *Pseudomonas*
1018 *aeruginosa* PA14 biofilms. Mol Microbiol 51:675-690.
1019 Gallie J, Libby E, Bertels F, Remigi P, Jendresen CB, Ferguson GC, Desprat N,
1020 Buffing MF, Sauer U, Beaumont HJ, et al. 2015. Bistability in a metabolic network
1021 underpins the de novo evolution of colony switching in Pseudomonas fluorescens.
1022 PLoS biology 13:e1002109. http://dx.doi.org/10.1371/journal.pbio.1002109
1023 Goymer P, Kahn SG, Malone JG, Gehrig SM, Spiers AJ, Rainey PB. 2006. Adaptive
1024 divergence in experimental populations of *Pseudomonas fluorescens*. II. Role of the
1025 GGDEF regulator WspR in evolution and development of the wrinkly spreader
1026 phenotype. Genetics 173:515-526. http://dx.doi.org/10.1534/genetics.106.055863
1027 Herron MD, Doebeli M. 2013. Parallel evolutionary dynamics of adaptive
1028 diversification in *Escherichia coli*. PLoS biology 11:e1001490.
1029 http://dx.doi.org/10.1371/journal.pbio.1001490
1030 Hickman JW, Tifrea DF, Harwood CS. 2005. A chemosensory system that regulates
1031 biofilm formation through modulation of cyclic diguanylate levels. Proceedings of the

1032 National Academy of Sciences of the United States of America 102:14422-14427.
1033 http://dx.doi.org/10.1073/pnas.0507170102
1034 Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, Bercot B, Petit
1035 E, Poulain J, Barnaud G, et al. 2013. Capturing the mutational landscape of the beta-
1036 lactamase TEM-1. Proceedings of the National Academy of Sciences of the United
1037 States of America 110:13067-13072. http://dx.doi.org/10.1073/pnas.1215206110
1038 Jahn LJ, Munck C, Ellabaan MMH, Sommer MOA. 2017. Adaptive Laboratory
1039 Evolution of Antibiotic Resistance Using Different Selection Regimes Lead to
1040 Similar Phenotypes and Genotypes. Front Microbiol 8:816.
1041 http://dx.doi.org/10.3389/fmicb.2017.00816
1042 Jerison ER, Desai MM. 2015. Genomic investigations of evolutionary dynamics and
1043 epistasis in microbial evolution experiments. Curr Opin Genet Dev 35:33-39.
1044 http://dx.doi.org/10.1016/j.gde.2015.08.008
1045 Jimenez JI, Xulvi-Brunet R, Campbell GW, Turk-MacLeod R, Chen IA. 2013.
1046 Comprehensive experimental fitness landscape and evolutionary network for small
1047 RNA. Proceedings of the National Academy of Sciences of the United States of
1048 America 110:14984-14989. http://dx.doi.org/10.1073/pnas.1307604110
1049 Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web
1050 portal for protein modeling, prediction and analysis. Nature protocols 10:845-858.
1051 http://dx.doi.org/10.1038/nprot.2015.053
1052 Kishimoto T, Iijima L, Tatsumi M, Ono N, Oyake A, Hashimoto T, Matsuo M,
1053 Okubo M, Suzuki S, Mori K, et al. 2010. Transition from positive to neutral in
1054 mutation fixation along with continuing rising fitness in thermal adaptive evolution.
1055 PLoS genetics 6:e1001164. http://dx.doi.org/10.1371/journal.pgen.1001164

1056 Knöppel A, Knopp M, Albrecht LM, Lundin E, Lustig U, Näsvall J, Andersson DI.
1057 2018. Genetic Adaptation to Growth Under Laboratory Conditions in *Escherichia coli*
1058 and *Salmonella enterica*. Front Microbiol 9:756.
1059 http://dx.doi.org/10.3389/fmicb.2018.00756
1060 Koskiniemi S, Sun S, Berg OG, Andersson DI. 2012. Selection-driven gene loss in
1061 bacteria. PLoS genetics 8:e1002787. http://dx.doi.org/10.1371/journal.pgen.1002787
1062 Kram KE, Geiger C, Ismail WM, Lee H, Tang H, Foster PL, Finkel SE. 2017.
1063 Adaptation of *Escherichia coli* to Long-Term Serial Passage in Complex Medium:
1064 Evidence of Parallel Evolution. mSystems 2.
1065 http://dx.doi.org/10.1128/mSystems.00192-16
1066 LaCroix RA, Sandberg TE, O'Brien EJ, Utrilla J, Ebrahim A, Guzman GI, Szubin R,
1067 Palsson BO, Feist AM. 2015. Use of adaptive laboratory evolution to discover key
1068 mutations enabling rapid growth of *Escherichia coli* K-12 MG1655 on glucose
1069 minimal medium. Appl Environ Microbiol 81:17-30.
1070 http://dx.doi.org/10.1128/AEM.02246-14
1071 Lambertsen L, Sternberg C, Molin S. 2004. Mini-Tn7 transposons for site-specific
1072 tagging of bacteria with fluorescent proteins. Environmental microbiology 6:726-732.
1073 http://dx.doi.org/10.1111/j.1462-2920.2004.00605.x
1074 Lee MC, Marx CJ. 2012. Repeated, selection-driven genome reduction of accessory
1075 genes in experimental populations. PLoS genetics 8:e1002651.
1076 http://dx.doi.org/10.1371/journal.pgen.1002651
1077 Lee VT, Matewish JM, Kessler JL, Hyodo M, Hayakawa Y, Lory S. 2007. A cyclic-
1078 di-GMP receptor required for bacterial exopolysaccharide production. Mol Microbiol
1079 65:1474-1484. http://dx.doi.org/10.1111/j.1365-2958.2007.05879.x

1080 Liang ZX. 2015. The expanding roles of c-di-GMP in the biosynthesis of
1081 exopolysaccharides and secondary metabolites. Nat Prod Rep 32:663-683.
1082 http://dx.doi.org/10.1039/c4np00086b
1083 Lind PA, Arvidsson L, Berg OG, Andersson DI. 2017a. Variation in mutational
1084 robustness between different proteins and the predictability of fitness effects.
1085 Molecular biology and evolution 34:408-418.
1086 http://dx.doi.org/10.1093/molbev/msw239
1087 Lind PA, Farr AD, Rainey PB. 2017b. Evolutionary convergence in experimental
1088 *Pseudomonas* populations. ISME J 11:589-600.
1089 http://dx.doi.org/10.1038/ismej.2016.157
1090 Lind PA, Farr AD, Rainey PB. 2015. Experimental evolution reveals hidden diversity
1091 in evolutionary pathways. eLife 4. http://dx.doi.org/10.7554/eLife.07074
1092 Lind PA, Libby E, Herzog J, Rainey PB. 2019. Predicting mutational routes to new
1093 adaptive phenotypes. eLife 8. http://dx.doi.org/10.7554/eLife.38822
1094 Long A, Liti G, Luptak A, Tenaillon O. 2015. Elucidating the molecular architecture
1095 of adaptation via evolve and resequence experiments. Nature reviews. Genetics
1096 16:567-582. http://dx.doi.org/10.1038/nrg3937
1097 Lukacisinova M, Bollenbach T. 2017. Toward a quantitative understanding of
1098 antibiotic resistance evolution. Curr Opin Biotechnol 46:90-97.
1099 http://dx.doi.org/10.1016/j.copbio.2017.02.013
1100 Lundin E, Tang PC, Guy L, Nasvall J, Andersson DI. 2017. Experimental
1101 determination and prediction of the fitness effects of random point mutations in the
1102 biosynthetic enzyme HisA. Molecular biology and evolution.
1103 http://dx.doi.org/10.1093/molbev/msx325

1104 McDonald MJ, Cooper TF, Beaumont HJE, Rainey PB. 2011. The distribution of
1105 fitness effects of new beneficial mutations in *Pseudomonas fluorescens*. Biology
1106 Letters 7:98-100. http://dx.doi.org/Doi 10.1098/Rsbl.2010.0547
1107 McDonald MJ, Gehrig SM, Meintjes PL, Zhang XX, Rainey PB. 2009. Adaptive
1108 divergence in experimental populations of *Pseudomonas fluorescens*. IV. Genetic
1109 constraints guide evolutionary trajectories in a parallel adaptive radiation. Genetics
1110 183:1041-1053. http://dx.doi.org/10.1534/genetics.109.107110
1111 McElroy KE, Hui JG, Woo JK, Luk AW, Webb JS, Kjelleberg S, Rice SA, Thomas
1112 T. 2014. Strain-specific parallel evolution drives short-term diversification during
1113 *Pseudomonas aeruginosa* biofilm formation. Proceedings of the National Academy of
1114 Sciences of the United States of America 111:E1419-1427.
1115 http://dx.doi.org/10.1073/pnas.1314340111
1116 Morgan JL, McNamara JT, Zimmer J. 2014. Mechanism of activation of bacterial
1117 cellulose synthase by cyclic di-GMP. Nat Struct Mol Biol 21:489-496.
1118 http://dx.doi.org/10.1038/nsmb.2803
1119 O'Neill AJ, Huovinen T, Fishwick CW, Chopra I. 2006. Molecular genetic and
1120 structural modeling studies of *Staphylococcus aureus* RNA polymerase and the
1121 fitness of rifampin resistance genotypes in relation to clinical prevalence. Antimicrob
1122 Agents Chemother 50:298-309. http://dx.doi.org/10.1128/AAC.50.1.298-309.2006
1123 Orgogozo V. 2015. Replaying the tape of life in the twenty-first century. Interface
1124 Focus 5:20150057. http://dx.doi.org/10.1098/rsfs.2015.0057
1125 Rainey PB, Travisano M. 1998. Adaptive radiation in a heterogeneous environment.
1126 Nature 394:69-72. http://dx.doi.org/10.1038/27900

1127 Römling U, Galperin MY, Gomelsky M. 2013. Cyclic di-GMP: the first 25 years of a
1128 universal bacterial second messenger. Microbiology and molecular biology reviews :
1129 MMBR 77:1-52. http://dx.doi.org/10.1128/MMBR.00043-12
1130 Sandberg TE, Pedersen M, LaCroix RA, Ebrahim A, Bonde M, Herrgard MJ, Palsson
1131 BO, Sommer M, Feist AM. 2014. Evolution of *Escherichia coli* to 42 degrees C and
1132 subsequent genetic engineering reveals adaptive mechanisms and novel mutations.
1133 Molecular biology and evolution 31:2647-2662.
1134 http://dx.doi.org/10.1093/molbev/msu209
1135 Sankar TS, Wastuwidyaningtyas BD, Dong Y, Lewis SA, Wang JD. 2016. The nature
1136 of mutations induced by replication-transcription collisions. Nature.
1137 http://dx.doi.org/10.1038/nature18316
1138 Schenk MF, Szendro IG, Krug J, de Visser JA. 2012. Quantifying the adaptive
1139 potential of an antibiotic resistance enzyme. PLoS genetics 8:e1002783.
1140 http://dx.doi.org/10.1371/journal.pgen.1002783
1141 Silby MW, Cerdeno-Tarraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston
1142 GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SA, et al. 2009. Genomic and
1143 genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*.
1144 Genome biology 10:R51. http://dx.doi.org/10.1186/gb-2009-10-5-r51
1145 Sleight SC, Orlic C, Schneider D, Lenski RE. 2008. Genetic basis of evolutionary
1146 adaptation by *Escherichia coli* to stressful cycles of freezing, thawing and growth.
1147 Genetics 180:431-443. http://dx.doi.org/10.1534/genetics.108.091330
1148 Sommer MOA, Munck C, Toft-Kehler RV, Andersson DI. 2017. Prediction of
1149 antibiotic resistance: time for a new preclinical paradigm? Nature reviews.
1150 Microbiology 15:689-696. http://dx.doi.org/10.1038/nrmicro.2017.75

1151 Spiers AJ, Bohannon J, Gehrig SM, Rainey PB. 2003. Biofilm formation at the air-
1152 liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an
1153 acetylated form of cellulose. Molecular microbiology 50:15-27.
1154 Spiers AJ, Kahn SG, Bohannon J, Travisano M, Rainey PB. 2002. Adaptive
1155 divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and
1156 phenotypic bases of wrinkly spreader fitness. Genetics 161:33-46.
1157 Spiers AJ, Rainey PB. 2005. The *Pseudomonas fluorescens* SBW25 wrinkly spreader
1158 biofilm requires attachment factor, cellulose fibre and LPS interactions to maintain
1159 strength and integrity. Microbiology 151:2829-2839.
1160 http://dx.doi.org/10.1099/mic.0.27984-0
1161 Steiner S, Lori C, Boehm A, Jenal U. 2013. Allosteric activation of exopolysaccharide
1162 synthesis through cyclic di-GMP-stimulated protein-protein interaction. EMBO J
1163 32:354-368. http://dx.doi.org/10.1038/emboj.2012.315
1164 Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD,
1165 Gaut BS. 2012. The molecular diversity of adaptive convergence. Science 335:457-
1166 461. http://dx.doi.org/10.1126/science.1212986
1167 Whiteside MD, Winsor GL, Laird MR, Brinkman FS. 2013. OrtholugeDB: a bacterial
1168 and archaeal orthology resource for improved comparative genomic analysis. Nucleic
1169 Acids Res 41:D366-376. http://dx.doi.org/10.1093/nar/gks1241
1170 Whitney JC, Whitfield GB, Marmont LS, Yip P, Neculai AM, Lobsanov YD,
1171 Robinson H, Ohman DE, Howell PL. 2015. Dimeric c-di-GMP is required for post-
1172 translational regulation of alginate production in *Pseudomonas aeruginosa*. J Biol
1173 Chem 290:12451-12462. http://dx.doi.org/10.1074/jbc.M115.645051
1174 Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ. 1999. Different
1175 trajectories of parallel evolution during viral adaptation. Science 285:422-424.

1176 Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FS. 2016. Enhanced
1177 annotations and features for comparing thousands of *Pseudomonas* genomes in the
1178 *Pseudomonas* genome database. Nucleic Acids Res 44:D646-653.
1179 http://dx.doi.org/10.1093/nar/gkv1227
1180

Figure 5 - source data. Mutations found after experimental evolution

Genome position CP000076.1	Type	Change	Gene positio	Gene locus	Gene symbol	Effect	Mutation found in SBW25
859290	Transition	g->a	-129	PFL_0740	recC	rrfB terminator/recCBD promoter upstream awsXRO	No
859290	Transition	g->a	-129	PFL_0740	recC	rrfB terminator/recCBD promoter upstream awsXRO	No
868744	Transversion	c->a	101	PFL_0743	awsX, yfiR	S34Y	No
868932	Transversion	t->g	289	PFL_0743	awsX, yfiR	Y97D	No
869088	Transversion	t->g	445	PFL_0743	awsX, yfiR	C149G	No
869281	Transversion	t->g	638	PFL_0743	awsX, yfiR	V213G	No
869281	Transversion	t->g	638	PFL_0743	awsX, yfiR	V213G	No
869281	Transversion	t->g	638	PFL_0743	awsX, yfiR	V213G	No
868832-868840	Deletion	Deletion 9 bp	189-197	PFL_0743	awsX, yfiR	Deletion F63-L65	No
868832-868840	Deletion	Deletion 9 bp	189-197	PFL_0743	awsX, yfiR	Deletion F63-L65	No
869040-869072	Deletion	Deletion 33 bp	397-429	PFL_0743	awsX, yfiR	Deletion V133-P143	No
869913	Transition	c->t	581	PFL_0744	awsR, yfiN	A194V	Yes
869994	Transition	a->g	662	PFL_0744	awsR, yfiN	D221G	D->A
870002	Transition	c->t	670	PFL_0744	awsR, yfiN	L226F	No
1301093	Transition	c->t	1982	PFL_1133	wspE	S661L	Yes
1302246	Transition	g->a	808	PFL_1134	wpsF	G270R	Yes
1302250	Transversion	t->g	812	PFL_1134	wspF	V271G	No
1302250	Transversion	t->g	812	PFL_1134	wspF	V271G	No
1302250	Transversion	t->g	812	PFL_1134	wspF	V271G	No
1302250	Transversion	t->g	812	PFL_1134	wspF	V271G	No
1302250	Transversion	t->g	812	PFL_1134	wspF	V271G	No
1302250	Transversion	t->g	812	PFL_1134	wspF	V271G	No
1302250	Transversion	t->g	812	PFL_1134	wspF	V271G	No
1302250	Transversion	t->g	812	PFL_1134	wspF	V271G	No

1302250	Transversion	t->g	812	PFL_1134	wspF	V271G	No
1302250	Transversion	t->g	812	PFL_1134	wspF	V271G	No
1302250	Transversion	t->g	812	PFL_1134	wspF	V271G	No
1302327	Transition	c->a	889	PFL_1134	wspF	Q297K	Q->R
1302327	Transition	c->a	889	PFL_1134	wspF	Q297K	Q->R
1301579-1301593	Duplication	Duplication 15 bp	141-155	PFL_1134	wspF	Duplication L51-I55	Del L51-I55
3548383	Transition	c->t	-98	PFL_3078		PFL_3078-PFL_3093 promoter	
3548383	Transition	c->t	-98	PFL_3078		PFL_3078-PFL_3093 promoter	
3548385	Transition	c->t	-96	PFL_3078		PFL_3078-PFL_3093 promoter	
6114402	Transversion	a->t	2237	PFL_5345	mwsR, morA	Q746L	No
6114696	Transition	a->g	2531	PFL_5345	mwsR, morA	H844R	No
6114940	Transition	g->a	2775	PFL_5345	mwsR, morA	M925I	Yes
6115236	Transversion	t->a	3071	PFL_5345	mwsR, morA	L1024Q	No
6115406	Transition	g->a	3241	PFL_5345	mwsR, morA	E1081K	Yes
6115406	Transition	g->a	3241	PFL_5345	mwsR, morA	E1081K	Yes
6115406	Transition	g->a	3241	PFL_5345	mwsR, morA	E1081K	Yes
6115406	Transition	g->a	3241	PFL_5345	mwsR, morA	E1081K	Yes
6115415	Transition	g->a	3250	PFL_5345	mwsR, morA	G1084R	G->S
6115107-6115130	Duplication	Duplication 24 bp	2942-2965	PFL_5345	mwsR, morA	P981-S988	No

Δ*wsp* Δ*aws* Δ*mws*							
86862-86873	Deletion	Deletion 12 bp	1389-1400	PFL_0087		S463R, deletion N464-Q466	
86876-86893	Deletion	Deletion 18 bp	1403-1420	PFL_0087		Deletion D470-S475	
86949-86978	Deletion	Deletion 30 bp	1476-1505	PFL_0087		Deletion Q494-L503	
86949-86978	Deletion	Deletion 30 bp	1476-1505	PFL_0087		Deletion Q494-L503	
86972-87172	Duplication	Duplication 201 bp	1499-1699	PFL_0087		Insertion 67 aa	
87014-87199	Duplication	Duplication 185 bp	1541-1726	PFL_0087		Duplication E544-P571	

Figure 7 - source data. Fitness assays

Invasion - selection coefficients

Wild type	0,003	0,002	-0,005	0,012	0,013	0,010	-0,039	0,021	-0,017
WspF V271G	0,363	0,330	0,303	0,318	0,319	0,340			
WspF Q297K	0,362	0,225	0,347	0,331	0,269	0,282			
WspE S661L	0,334	0,349	0,377	0,362	0,335	0,391			
AwsX del V133-P143	0,300	0,219	0,358	0,223	0,271	0,261			
AwsR A194V	0,308	0,351	0,280	0,262	0,292	0,262			
Promoter AwsXRO	0,412	0,526	0,476	0,348	0,336	0,376			
MwsR M925I	0,392	0,394	0,432	0,346	0,341	0,373			
MwsR E1081K	0,322	0,420	0,357	0,326	0,321	0,283			
PFL_3078 Promoter	0,313	0,351	0,320	0,321	0,334	0,326			
PFL_0087 D470-S475	0,390	0,407	0,433	0,393	0,417	0,429			
PFL_0087 Q494-L503	0,428	0,378	0,383	0,419	0,453	0,433			
WspA del T293-E299	0,435	0,439	0,440	0,444	0,429	0,416			

Competition - selection coefficients

Wild type	-0,230	-0,207	-0,196	-0,175				
WspF V271G	0,033	-0,021	-0,011	-0,029	-0,014	0,027	-0,015	0,010
WspF Q297K	0,015	0,017	-0,009	-0,010				
WspE S661L	-0,008	0,001	0,002	0,050				
AwsX ΔV133-P143	-0,030	-0,082	-0,088	-0,058				
AwsR A194V	-0,113	-0,079	0,008	-0,020				
Promoter AwsXRO	0,041	0,009	0,034	0,039				
MwsR M925I	-0,006	-0,007	-0,011	0,003				
MwsR E1081K	0,085	0,072	0,036	0,050				
PFL_3078 Promoter	-0,066	-0,146	-0,088	-0,102				
PFL_0087 D470-S475	-0,055	-0,056	-0,069	-0,089				
PFL_0087 Q494-L503	-0,056	-0,050	-0,030	-0,026				
WspA del T293-E299	-0,145	-0,068	-0,062	-0,049				

Figure 8 - source data. Fitness assays for del pel mutants

Invasion - selection coefficients

Δ*pel* WspF V271G	0,212	0,253	0,227	0,199	0,197	0,201
Δ*pel* WspF Q297K	0,210	0,162	0,173	0,176	0,212	0,223
Δ*pel* AwsX del V133-P143	0,311	0,222	0,184	0,202	0,203	0,178
Δ*pel* MwsR E1081K	0,225	0,209	0,248	0,204	0,185	0,205
Δ*pel* PFL_0087 D470-S475	0,223	0,244	0,206	0,222	0,259	0,257

Competition - selection coefficients

Δ*pel* WspF V271G	-0,171	-0,177	-0,206	-0,174
Δ*pel* WspF Q297K	-0,041	-0,114	-0,104	-0,069
Δ*pel* AwsX del V133-P143	-0,171	-0,104	-0,200	-0,151
Δ*pel* MwsR E1081K	-0,042	-0,025	-0,112	-0,072
Δ*pel* PFL_0087 D470-S475	-0,179	-0,255	-0,179	-0,237

Table S1. Oligonucleotide primers used for PCR, sequencing and genetic reconst

Cloning	
pEX18Gm_MCS_F	tgttgtgtggaattgtgag
pEX18Gm_MCS_R	ctgcaaggcgattaagttg
***wsp* operon**	
PFL_1133_EcoRI_F	agttgctggcggagaaaac
WspE_1550F	ttcccgctggcccatatcga
WspE_2119F	atcaccgacatcgacatgc
WspE_2213F	gtcctacaaggaccgtga
WspF_154R	caccggcatgatcaggtccat
WspR_70R	gcttcgccgatcatggcctg
WspR_398F	tggtggcgcgcattcgcta
WspR_553R	ccactccagctccaggta
WspR_620R	gtaggtcttgaagtagtcg
PFL_1128_F_EcoRI	tagGAATTCcaagaaggcccgcaagtac
PFL_1128_UR_SOE	cgtatcctgctagattccgaatgagtcagtcctcgaggccat
PFL_1135_DF_SOE	cattcggaatctagcaggatacggagtaaccgcctcagcttc
PFL_1136_F_SalI	catGTCGACgaatcactgattcacggcag
***aws* operon**	
RecC_F_HindIII	tagaagcttctgataccgcccaagagttc
RecC_R_KpnI	catggtaccccagtcggctcgatatacct
recD_F_EcoRI	acggaattcccactacctgaatgtactg
PFL_0743_R_SOE	cgttagaaatcggctaggcttcagagaaaggccggaaacatg
PFL_0743-SalI_F	tgtgtgctgtcgaccattc
awsR_182R	agctgatggagcgggcgatca
awsR_677F	cgacttcaacgccctgct
PFL_0745_F_SOE	tgaagcctagccgatttctaacgatcgtggtgatcgccgact
awsO_DR	ttacatccgcgaggtgac
PFL_0746_F_HindIII	tgcaagcttcttcttcatcaccccgcaa
mwsR	
PFL_5345_UF_EcoRI	actgaattcatggcttccagcttggcct
PFL_5345_UR_SOE	cagtcgattacgtactatagcaggctgtgatcctgcggtttg
mwsR_HindIII_F	gttaagcttagacccagctgttcctgttc
mwsR_2144F	gccgcgacatcagccagca
PFL_5345_SeqF	gaaaaggacctgcgcatg
PFL_5345_SeqR	gaactgcttgaggtagttc
mwsR_3599R	gaaggtgcggtcgatcttca
mwsR_SalI_R	aggccgtcgacgaaggtgc
PFL_5345_DF_SOE	ctgctatagtacgtaatcgactgcatcccctgcttgcgtaat
PFL_5346_R_HindIII	gtcaagctttcagcgatttcgcggaaac
PFL_0087	
PFL_0087_KpnI_F	catggtaccgaagccgttgcgacgaatc

PFL_0087_seqF	aagcaaccatttcgcctgt
PFL_0087_seqR	tgccatgttgttgccattg
PFL_0087_HindIII_R	tagaagcttggctcctcaaccagcttc

PFL_3078-3093

PFL_3077_KpnI_F	catggtaccgcgaaagtcccggttgaag
PFL_3078_HindIII_R	tagaagcttgttggccatctcgttcatg

pelABCDEFG

PFL_2971_F_EcoRI	ctggaattcagcgagtactacctggactt
PFL_2972_DelCh	caaggccaatgcggtaaaca
PFL_2972_UR_SOE	attgggccctctatgtcgacatcaactcactcttcagtagatcaatct
PFL_2978_DF_SOE	gatgtcgacatagagggcccaatttgaggagcatcggcaagc
PFL_2979_R_HindIII	agtgcccaagagcagaagc