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). 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; Knoppel, 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).

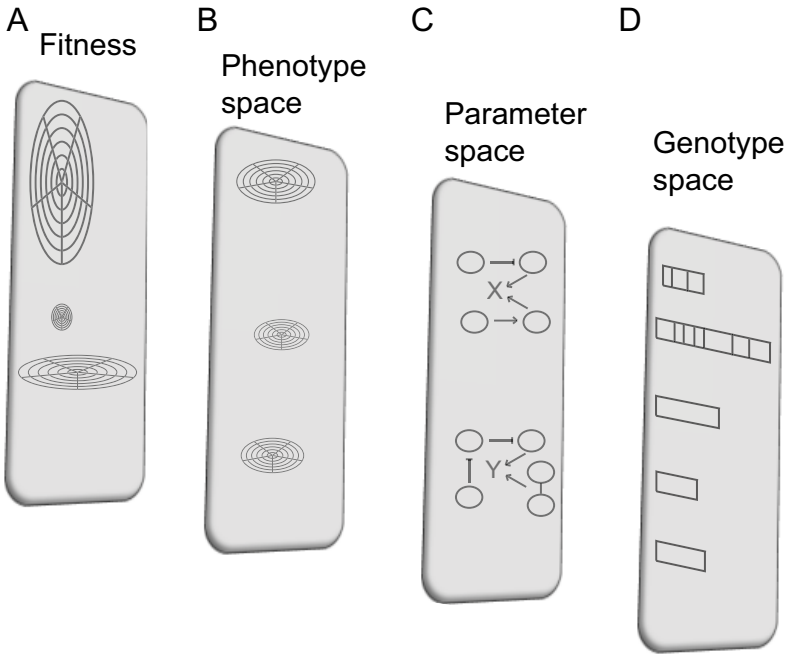


Figure 1. Prediction of the adaptive outcomes of experimental evolution requires understanding of how genotypes map to phenotypes and fitness. The differing capacity of genes to translate genotypic variation into phenotypic variation and

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 space**. 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 difference 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

97 with uniform rates. Alternatively experimental data might be incorporated into models
98 of mutation rates to improve predictions.
99
100
101 There are several problems that need to be solved to develop a model system for true
102 testing of predictive ability. In some cases adaptive mutations are highly strain
103 specific, so that for example adaptation of different strains to a specific environment
104 will produce different results. In some cases this is due to the long history of sub-
105 culturing under laboratory conditions combined with rounds of mutagenesis that has
106 caused, for example, many *E. coli* and *Salmonella* strains to accumulate diverse
107 mutations, some of which are rapidly compensated by secondary mutations restoring
108 fitness (Barrick, et al. 2009; Tenaillon, et al. 2012; Knoppel, et al. 2018). Thus, in
109 many cases conclusions from one strain cannot be extended to another because of
110 differences in their genotype-to-phenotype maps (Figure 1C).
111
112 Another problem for testing predictability is that in many cases it is not possible to
113 design an experiment where one specific selective pressure is dominant. For example
114 experiments with intended adaptation to high temperature (Tenaillon, et al. 2012) or
115 freeze-thaw-growth cycles (Sleight, et al. 2008) result in similar mutations in *uspA*,
116 which may indicate adaptation to the medium used (Knoppel, et al. 2018) or generally
117 stressful conditions. Relatively minor changes in environmental conditions can also
118 results in divergent mutational patterns (Deatherage, et al. 2017). This means that the
119 range of possible adaptive phenotypes cannot be defined beforehand (Figure 1B) and
120 that in many cases the phenotypes that solve the intended selective problem are
121 outcompeted by other phenotypes with increased fitness (Figure 1A).
122
123 A highly specific selective pressure can be applied by selection for antitbiotic
124 resistance and mutation targets are often highly conserved between different strains
125 and species and between laboratory and natural populations (O'Neill, et al. 2006;
126 Schenk, et al. 2012; Brandis, et al. 2015; Jahn, et al. 2017; Lukacisinova and
127 Bollenbach 2017; Sommer, et al. 2017). However resistance phenotypes are often
128 explained solely by the molecular phenotype of a single protein and no alternative
129 pathways to resistance are known resulting in a relatively simple parameter and
130 genotype space (Figure 1C, Figure 1D). Thus the prediction will be identical for all

131 species and it cannot provide a test of prediction from general principles. In many
132 cases mutants isolated after selection for high-level antibiotic resistance also lacks the
133 complexity that is inherent to many phenotypic traits where the genotype-to-
134 phenotype map involves a large number of functional interactions and complex
135 regulation (Figure 1C). This complexity comes to light in that adaptive mutations to
136 new environments are commonly found in global regulators of gene expression such
137 as genes involved in the stringent response, DNA binding proteins, supercoiling and
138 core genes for RNA and protein synthesis (Barrick, et al. 2009; Conrad, et al. 2009;
139 Kishimoto, et al. 2010; Tenaillon, et al. 2012; Herron and Doebeli 2013; Sandberg, et
140 al. 2014; LaCroix, et al. 2015; Deatherage, et al. 2017). The physiological effects of
141 these mutations are diverse, sometimes affecting the expression of hundreds of genes
142 making the elucidation of the molecular underpinnings of the adaptive phenotype
143 (Figure 1C, 1D) extremely complex and thus difficult to use for predictive modeling.
144
145 The wrinkly spreader model in *P. fluorescens* SBW25 (hereafter SBW25) is one of
146 the best-characterized experimental evolution systems and has several properties that
147 could make it possible to extend knowledge and principles from this species to related
148 species (Rainey and Travisano 1998; Spiers, et al. 2002; Spiers, et al. 2003; Spiers
149 and Rainey 2005; Goymer, et al. 2006; Bantinaki, et al. 2007; McDonald, et al. 2009;
150 Silby, et al. 2009; Ferguson, et al. 2013; Lind, et al. 2015, 2017b; Lind, et al. 2018).
151 When the wild type SBW25 is placed into a static growth tube the oxygen in the
152 medium is rapidly consumed by growing bacteria (Figure 2A). However oxygen
153 levels at the surface are high and mutants that are able to colonize the air-liquid
154 interface have a major growth advantage and rapidly increase in frequency (Figure
155 2A). Several phenotypic solutions to air-liquid interface colonization, all involving
156 increased cell-cell adhesion, have been described and are distinguishable by their
157 colony morphology on agar plates (Figure 2A, 2B) (Rainey and Travisano 1998;
158 Ferguson, et al. 2013; Lind, et al. 2017b). The most successful of these is the Wrinkly
159 Spreader (WS) (Ferguson, et al. 2013; Lind, et al. 2017b) that overproduces a
160 cellulosic polymer that is the main structural component of the mat at the air-liquid
161 interface (Spiers, et al. 2002; Spiers, et al. 2003). The WS phenotype is caused by
162 mutational activation of c-di-GMP production by a diguanylate cyclase (DGC)
163 (Figure 2C) (Goymer, et al. 2006). While many different DGCs can be activated to
164 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. 2018) 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. 2018).

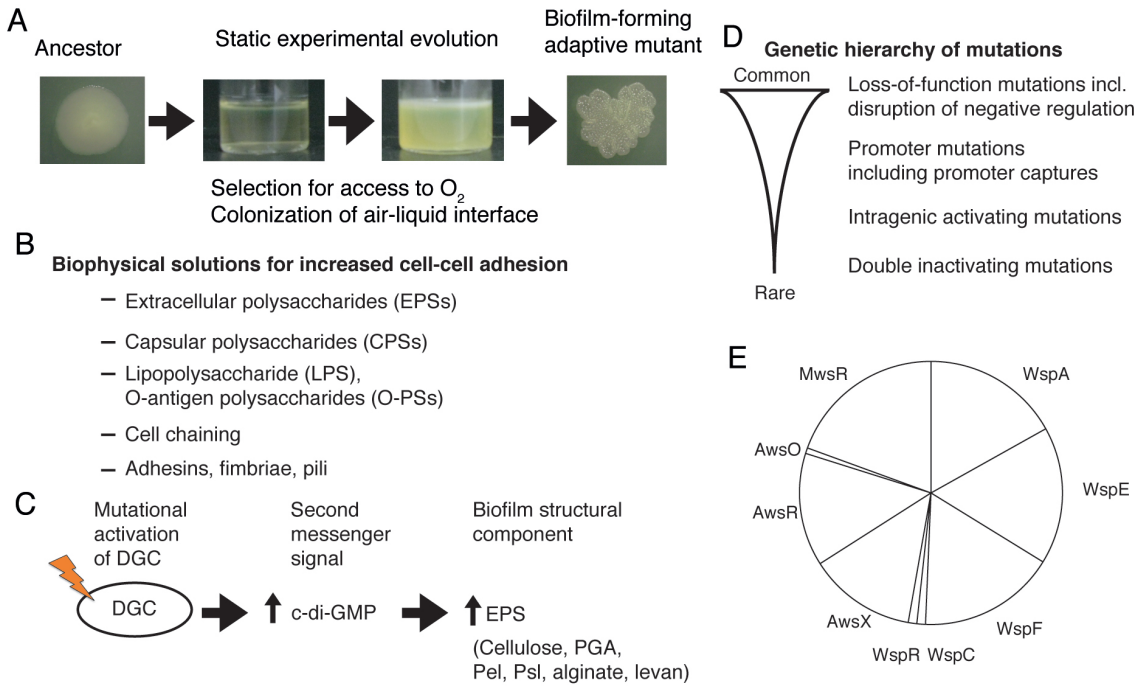


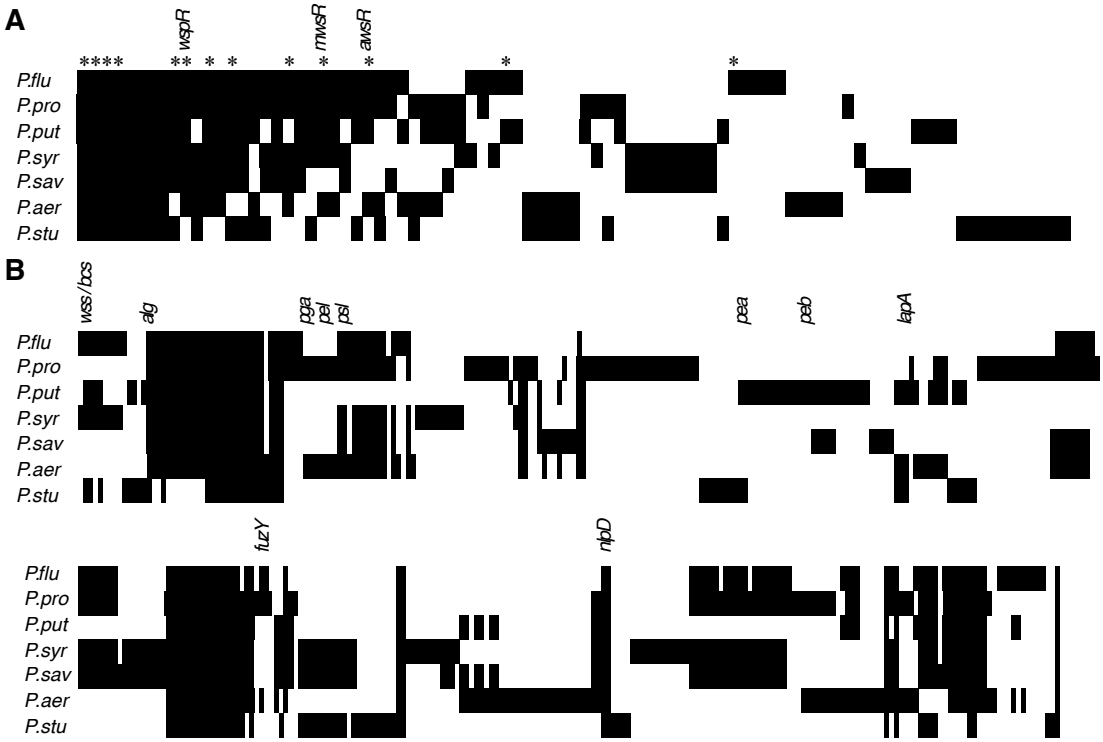
Figure 2. The *Pseudomonas fluorescens* SBW25 "wrinkly spreader" model system has several properties that could allow its extension to other species (**A**) The ancestral strain that has smooth colony morphology on agar plate is inoculated into static growth tubes and incubated for several days. Depletion 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**) There are many potential solutions to colonization of the the air-liquid interface and in *Pseudomonas fluorescens* at least four alternative distinct phenotypes are selected for

185 at the surface, but they all have lower fitness than the WS type (Beaumont, et al.
186 2009; Ferguson, et al. 2013; Gallie, et al. 2015; Lind, et al. 2017b). The SBW25
187 system is of intermediate complexity in that it allows isolation of a phenotypic subset
188 of adaptive mutants, those that are selected at the air-liquid interface and produce a
189 difference in colony morphology, rather than all mutants that increase fitness.
190 However there is also a large diversity of different genetic and phenotypic solutions to
191 the dominant adaptive challenge, which makes it of greater complexity than systems
192 based on single genes, which is often the case for experimental systems using strong
193 selection for antibiotic resistance. (**C**) The mutational causes of WS in wild type
194 populations are found in either of three loci, *wspABCDEFR*, *awsXRO* and *mwsR,*
195 which all encode diguanylate cyclases that produce the second messenger c-di-GMP
196 (McDonald, et al. 2009) that is a conserved signal for EPS production and biofilm
197 formation in many bacterial species. In SBW25 the primary EPS used is a cellulosic
198 polymer. (**D**) While these three pathways account for >98% of WS mutants in the
199 wild type, 13 additional rare pathways were found when the common ones are
200 genetically deleted (Lind, et al. 2015). The large differences in mutation rates to WS
201 for the different pathways are explained mainly by their different capacities to
202 translate genotypic variation into phenotypic variation, i.e. mutational target size, with
203 the three common pathways being subject to negative regulation, which when
204 disrupted results in overproduction of c-di-GMP and the WS phenotype (Lind, et al.
205 2015). The alternative phenotypic solutions are also caused by inactivating mutations
206 occurring at high rates (Ferguson, et al. 2013; Lind, et al. 2017b). The pathways to
207 WS of intermediate frequency are activated by mutations to promoter regions,
208 including promoter captures and the most rare are those that require specific
209 activating mutations in the DGC or double mutations in two negative regulators. The
210 genetic underpinnings of several adaptive phenotypes have been elucidated providing
211 a mechanistic understanding of the effects of mutations and why they are adaptive.
212 (**E**) The three main pathways (Wsp, Aws, Mws) to WS are particularly well-
213 understood allowing formulation of mathematical models of the molecular networks
214 and prediction of the relative rates of use of the different pathways and genes (Lind, et
215 al. 2018).
216
217 This study makes initial forecasts of phenotypic and genetic evolutionary outcomes
218 after static experimental evolution for six *Pseudomonas* species based mainly on data

219 from SBW25 (McDonald, et al. 2009; McDonald, et al. 2011; Ferguson, et al. 2013;
220 Lind, et al. 2015, 2017b). Predictions of evolutionary outcomes are then
221 experimentally tested for the closely related species, *P. protegens* Pf-5 (hereafter Pf-
222 5) with a highly conserved genetic repertoire of DGCs but that lack the structural
223 component used for biofilm formation in SBW25. Results show that phenotypes,
224 order of pathways used and types of mutations can be predicted and that forecasts are
225 robust to changes in environmental conditions.
226

227 ## Results

228 Six *Pseudomonas* species (Figure 3 legend) were chosen based on phylogenetic
229 diversity and their complement of DGCs and EPSs for a first round of predictions.
230 These species encode from none to all three of the main DGCs used in SBW25 and
231 only three species contain genes related to cellulose biosynthesis, the main EPS used
232 in SBW25 (Figure 3).



233

234 **Fig 3. Diversity of DGCs and biofilm-related genes for seven *Pseudomonas***
235 **species** (*P. fluorescens* SBW25, P. *protegens* Pf-5, *P. putida* KT2440, *P. syringae* pv.
236 tomato DC3000, *P. savastanoi* pv. phaseolicola 1448A, *P. aeruginosa* PAO1, *P.*
237 *stutzeri* ATCC 17588) **(A)** The seven species encode 251 putative DGCs, divided into
238 87 different homolog groups of which 8 are present in all genomes. WS mutations in
239 SBW25 have been found affecting 13 of these DGCs (marked with *) with an

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

274 LPS modification (fuzzy spreaders FS) and cell chaining (CC) (Spiers, et al. 2002;
275 Ferguson, et al. 2013; Lind, et al. 2017b). However wrinkly spreaders that form
276 cellulose-based mats are superior and rapidly outcompete all other types (Ferguson, et
277 al. 2013; Lind, et al. 2017b). Based on the limited data available it is predicted that
278 cellulose-based biofilms are superior in other species as well and that they will be the
279 primary structural solution when available as for *P. syringae*, *P. putida* and *P.*
280 *stutzeri*. For the three species lacking genes for cellulose biosynthesis, other EPSs are
281 predicted to be used. Based on studies of *P. aeruginosa* the primary EPS required for
282 pellicle formation at the air-liquid interface in this species is Pel, encoded by the
283 *pelABCDEFG* operon, which is also present in the Pf-5 genome and is predicted to be
284 the primary phenotypic solution for these species. The genome of *P. savastanoi* lacks
285 genes for biosynthesis of cellulose and Pel as well as other EPSs that are known to be
286 able to support mat-formation, such as PGA and there is not sufficient data at this
287 point to make a prediction of which one is likely to be the primary phenotypic
288 solution.
289
290 Overexpression of EPSs used for mat-formation at the air-liquid interface is in
291 SBW25 and *P. aeruginosa* linked to mutations increasing c-di-GMP production rather
292 than mutations in the promoters of or genes in the EPS operons themselves. This can
293 be explained by the role of post-translation regulation by c-di-GMP in the production
294 of cellulose, Pel, PGA and alginate (Lee, et al. 2007; Romling, et al. 2013; Steiner, et
295 al. 2013; Morgan, et al. 2014; Liang 2015; Whitney, et al. 2015). In these cases
296 transcriptional upregulation alone is not likely to cause overproduction because of
297 lack of a c-di-GMP signal. Possibly there is also an additional benefit to using
298 activation of the c-di-GMP network in that it reduces motility, which is not needed
299 when established at the air-liquid interface and which consumes large amount of
300 energy to sustain and thus is likely to be selected against (Koskiniemi, et al. 2012; Lee
301 and Marx 2012).
302
303 **Prediction of types of mutations**
304 Disabling mutations are expected to be more common than enabling mutations and
305 therefore the prediction is that most mutations will be in genes where loss-of-function
306 mutations produce an adaptive phenotype (Fig 2D) (Lind, et al. 2015). This is the case
307 for the large majority of mutations in SBW25 including those activating main DGCs

308 WspR, AwsR and MwsR, which are all under negative regulation, as well as
309 disruption of the genes underpinning the FS phenotype (*fuzY*, PFLU0478) and CC
310 phenotype (*nlpD*, PFLU1301) (McDonald, et al. 2009; Ferguson, et al. 2013; Lind, et
311 al. 2017b). Next in the hierarchy of mutations are promoter mutations, increasing
312 transcription, and promoter capture events (Lind, et al. 2015). Less common are
313 intragenic activating mutations that enable a particular function by for example
314 increase in catalytic activity or strengthening of interactions to another molecule or
315 another domain of the same protein (Lind, et al. 2015). Gene duplications occur at a
316 high rate and clearly have the ability to increase gene expression of DGCs, but they
317 have not yet been found to cause WS in SBW25, possibly because a two-fold increase
318 in gene expression is insufficient.
319
320 **Prediction of pathways used**
321 There are at least 16 different pathways to the WS phenotype in SBW25 with similar
322 fitness, but they are used at frequencies that vary over several orders of magnitude
323 based on the differing capacity to translate phenotypic variation into phenotypic
324 variation (Figure 2D) (Lind, et al. 2015). Mutations in three pathways, Wsp, Aws, and
325 Mws account for >98% of WS mutations and based on a detailed understanding of the
326 molecular functions of the genes involved of each pathways mathematical models
327 predicting at which relative rates the pathways should be used were constructed
328 (Figure 2E) (Lind, et al. 2018). The prediction results varies depending on the rates of
329 disabling and enabling mutations, but if it is assumed that disabling mutations are at
330 least an order of magnitude more common than enabling mutations the models predict
331 that Wsp will account for about 53%, Aws 28% and Mws 19% of the WS mutations
332 (Figure 2E) (Lind, et al. 2018).
333
334 Less common promoter mutations will also appear at rates at least a magnitude lower,
335 but which DGCs that will be transcriptionally activated cannot be easily predicted
336 except for assuming it will be homologs of the ones used in SBW25. These DGCs
337 must be catalytically active and also be localized to the membrane (Farr, et al. 2017).
338 Possibly the subset of DGCs that are primarily activated by mutations to their
339 promoters is mainly determined by mutation rate and a higher mutation rate might be
340 caused by higher transcription and also influenced by gene direction (Sankar, et al.
341 2016). Most promoter capture deletion events are less than five kilobases (Lind, et al.

342 2015) in size and the lack of an alternative promoter relatively close upstream to the
343 DGC is likely to rule out these DGCs. The DGCs that can be activated by intragenic
344 activating mutations cannot now be predicted beyond the simple prediction that these
345 are the same genes as in SBW25 (Lind, et al. 2015).
346
347 **Prediction of mutated genes**
348 In addition to predicting the relative rates of the three main pathways, the previously
349 described mathematical model can also predict which proteins are likely to be
350 mutated (Lind, et al. 2018). High rates of WS mutations are predicted for WspF,
351 WspA, WspE, AwsX and AwsR and MwsR (Figure 2E). A significantly lower rate of
352 enabling mutations is also predicted to occur in WspC, WspR and AwsO (Figure 2E).
353 Despite the simplicity of null model it closely predicted the mutational targets in
354 SBW25 with equal rates for WspF, WspA and WspE and rare mutations in WspC and
355 WspR, suggesting that it is a useful null model also for other species (Lind, et al.
356 2018).
357
358 **Prediction of specific mutational targets and effects of mutations**
359 The level of parallelism at the nucleotide level between species is expected to be
360 dependent both on the number of possible mutations to WS and the degree of
361 functional conservation of the proteins involved that define the genotype-to-
362 phenotype map. Mutational hot spots are also expected to contribute to parallelism
363 when they are conserved, but reduce parallelism when they are not. Based on previous
364 analysis of patterns of mutations in SBW25 (McDonald, et al. 2009; McDonald, et al.
365 2011; Lind, et al. 2018) and homology modeling of protein structure using Phyre2
366 (Kelley, et al. 2015) regions expected to be mutated were predicted and the likely
367 molecular consequences of different mutations suggested (Figure 4).
368

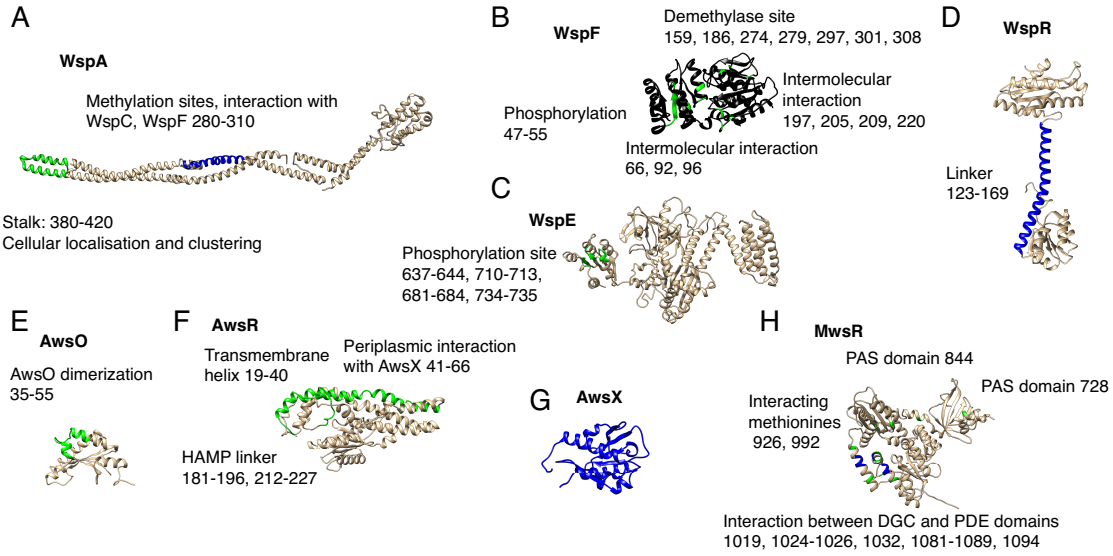


Figure 4. Predicted mutational targets and proposed molecular effects. Black represents any inactivating mutation including frame shifts, blue represents in frame inactivating mutations, green represents amino acid substitutions. Numbers in brackets refers to amino acid residue numbers in SBW25 (**A**)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

Prediction of fitness effects of WS mutations

While conservation of relative fitness of different phenotypic variants might be expected there is no clear reason to expect that the relative fitness of different DCG pathways and mutations will be conserved between species. Despite this difficulty

393 there might be a way forward to predict the relative fitness of a large range of
394 mutations with limited experimental data. The distribution of fitness effects of new
395 mutations have been found to be bimodal for a large number of genes with different
396 functions with one mode close to neutrality and one corresponding to a complete loss
397 of a particular molecular function (Jacquier, et al. 2013; Jimenez, et al. 2013;
398 Firnberg, et al. 2014; Lind, et al. 2017a; Lundin, et al. 2017). Given that mutations
399 that allow colonization of the air-liquid interface have large phenotypic effects and
400 are believed to also have large effects on molecular function, often a complete
401 disruption of an interaction, adaptive mutations in the same region of a protein are
402 likely to have similar fitness effects. Thus, an approximation of the distribution of
403 fitness effects could be possible with relative few mutations for each gene. This is
404 supported by the relatively small number of WS mutants in SBW25 that have been
405 characterized with sensitive fitness assays and where mutations in the same gene
406 typically have similar fitness effects (Lind, et al. 2015; Lind, et al. 2018). If this
407 assumption is true the distribution of beneficial fitness effects is not continuous and
408 the most advantageous mutations are not predicted to be equally spread between
409 pathways or genes. Thus the prediction would be that mutants isolated after
410 experimental evolution were concentrated to certain genes even if the mutational rate
411 is similar so that although the prediction from the null model is equal number of
412 mutations for WspA, WspE and WspF such distribution is unlikely to be found. While
413 the mutation rates to WS for the three genes are similar in SBW25, WspA mutants are
414 rarely found after experimental evolution due to their lower fitness (McDonald, et al.
415 2009; Lind, et al. 2018). There is however no reason to expect that the relative fitness
416 of mutations in different genes or pathways will be conserved between species.
417
418 Inactivating mutations in *fuzY* and *nlpD* producing the alternative adaptive
419 phenotypes based on LPS modification or cell chaining were also found to have
420 similar fitness (Ferguson, et al. 2013; Farr 2015). Possibly there are other genes that
421 can be mutated with similar phenotypes, but that those mutants have lower fitness and
422 are outcompeted in SBW25. If relative fitness is not conserved between species this
423 could lead to high convergence on the phenotypic level but with completely different
424 genetic bases.
425
426 **Experimental test of forecasts in *P. protegens* Pf-5**

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

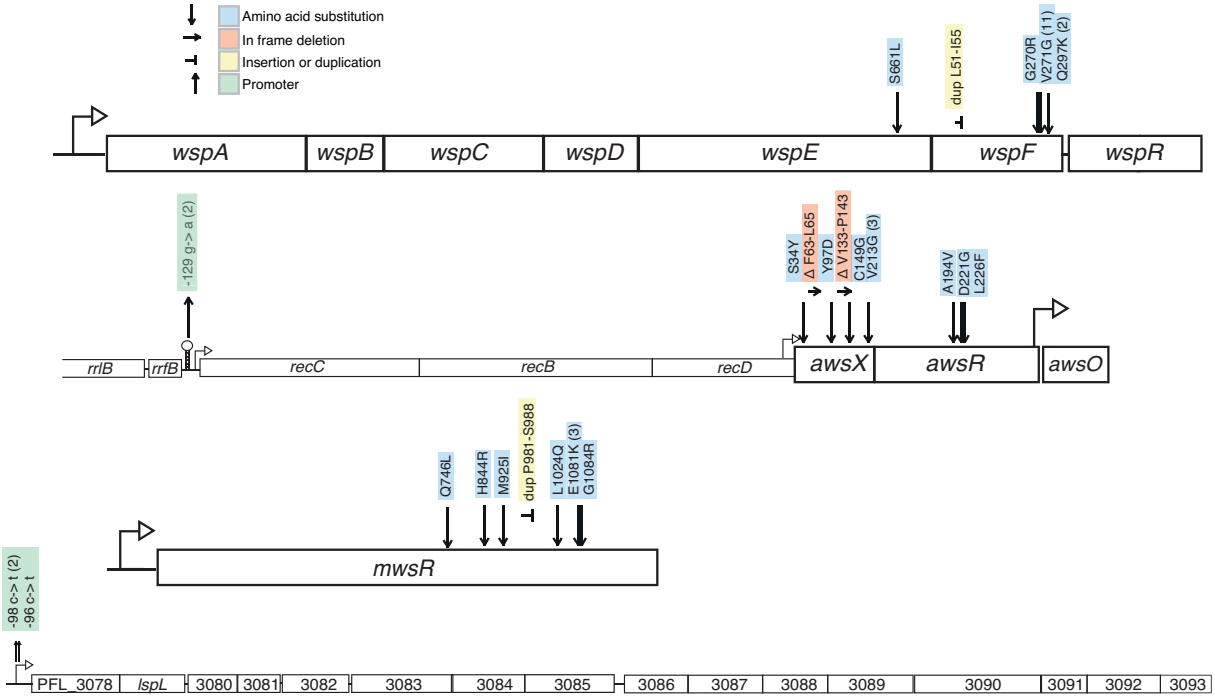


Figure 5. 43 independent mutants of *Pseudomonas protegens* Pf-5 were isolated after experimental evolution based on their divergent colony morphology and mutations were identified in four operons. Numbers in brackets are the number of independent mutants found. Details are available in Figure 5 – source data.

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. 2018) appeared absent, demonstrating how mutation rate differences can skew evolutionary outcomes even for closely related species with similar genetic architecture.

In total 60 wells were inoculated and subjected to experimental evolution for five days after which air-liquid interface colonization was observed for the majority of the wells. Mutants with clearly visible changes in colony morphology were isolated from 43 wells. Typically a single type of divergent colonies was observed and one colony for each well was selected for further characterization at random based on a pre-determined position on the agar plate. Representative mutations were reconstructed using an allelic exchange protocol to determine that the mutations are the sole cause of the air-liquid interface colonization and colony phenotypes and to exclude the influence of secondary mutations (Figure 6A) before further characterization.

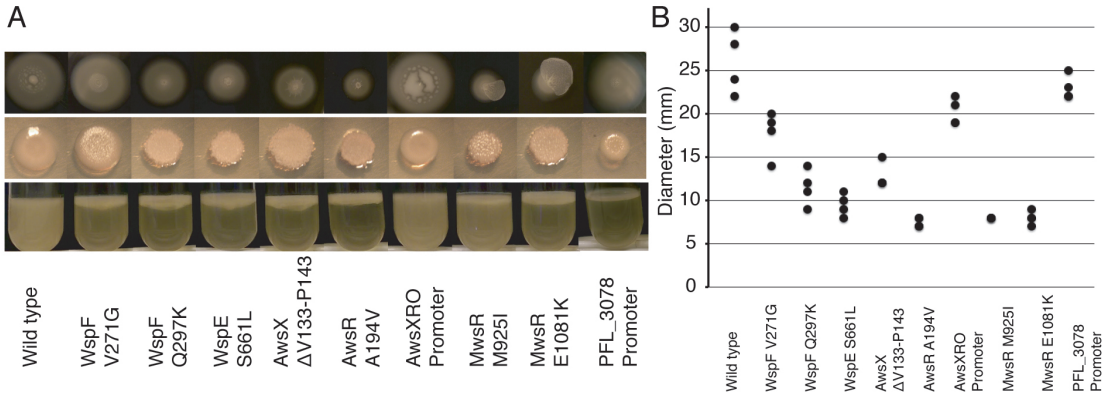


Figure 6. Phenotypic characterization of reconstructed mutants in WspF (V271G, Q297K), WspE (S661L), AwsX (deletion V133-P143), AwsR (A194V), MwsR (M925I, E1081K) and upstream *awsXRO* (*recC* -129 g->a) and PFL_3078 (-98 c->t) (**A**) Motility, colony morphology and air-liquid interface colonization (**B**) Motility assay. As expected if the c-di-GMP network is activated motility was reduced for most mutants with the exception of PFL_3078 that is not expected to have increased c-di-GMP levels.

The lack of cellulose biosynthetic genes also shows that these ecotypes can evolve by different phenotypes than in SBW25. Increased Congo Red binding ability suggests than alternative EPS components are used. Two clearly different phenotypes were observed with one very similar to the original WS types in SBW25 with a clear motility defect and mutations in the Wsp, Aws and Mws pathways (Figure 6A, 6B). The other type was less wrinkly, had similar motility as the wild type and promoter mutations upstream of the PFL_3078-3093 operon (Figure 6A, 6B).

503

504 Two types of fitness assays were performed, similarly as previously described (Lind,
505 et al. 2015) to measure differences in fitness between the different DGC mutants and
506 the alternative phenotypic solution with the mutation upstream of PFL_3078. The first
507 assay measures "invasion fitness" where the mutant is allowed to invade a wild type
508 population from an initial frequency of 1%. This confirms that the mutations are
509 adaptive and that mutants can colonize the air-liquid interface. The invasion assays
510 showed that all reconstructed mutants could rapidly invade an ancestral wild type
511 population (Figure 7A, Figure 7 – source data). Although there were significant
512 differences between selection coefficients of the mutants (one-way ANOVA $p <$
513 0.0001), no mutant was significantly different from the most common mutant (WspF
514 V271G, two-tailed t-test $P > 0.01$).
515

516 The second fitness assay measures "competition fitness" and here each mutant is
517 instead mixed 1:1 with the most common WS type (WspF V271G) at the start of the
518 competition. The competition assay showed that the ancestral wild type was rapidly
519 outcompeted by the mutants also at a 1:1 initial ratio (Figure 7B). There was
520 significant variation in fitness between the WS mutants (one-way ANOVA $P <$
521 0.0001) and the AwsX had significantly lower selection coefficient (two-tailed t-test P
522 < 0.009) compared to the reference WspF V271G and one of the MwsR mutants
523 (E1081K) had significantly higher selection coefficient (two-tailed t-test $P < 0.003$).
524 The alternative phenotypic solution used by the PFL_3078 promoter mutant resulted
525 in the lowest fitness ($s = -0.1$, two-tailed t-test $p < 0.005$) meaning that it is expected
526 to be rapidly outcompeted by the WS mutants (Figure 7B).

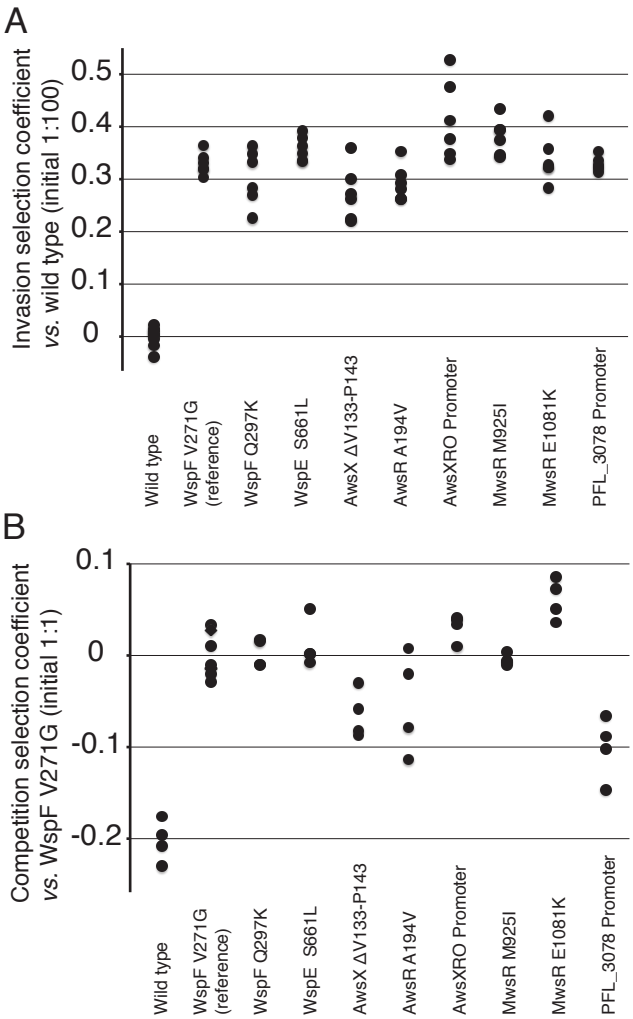


Figure 7. Fitness of reconstructed *P. protegens* Pf-5 WS mutants was measured in pairwise competitions. **(A)** Invasion fitness was measured relative a dominant ancestral wild type strain with a 1:100 initial ratio to show that mutations were adaptive and that they can increase from rare to colonize 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.

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 case at the phenotypic level as its genome does not encode genes for cellulose biosynthesis. Given that production of Pel exopolysaccharide has been shown to be induced by mutations in *wspF* in *P. aeruginosa* (Hickman, et al. 2005) and that Pel in this species is required for pellicle formation under static growth (Friedman and

543 Kolter 2004) this was predicted to be the main structural component used by Pf-5. To
544 test this prediction the *pel* operon (PFL_2972-PFL_2978) was deleted from Pf-5 and
545 combined with previously characterized WS mutations and fitness was measured.
546 Both invasion fitness (Figure 8A) and competition fitness (Figure 8B) was
547 significantly lower (two-tailed t-tests $p < 0.01$) compared to isogenic strains with an
548 intact *pel* operon (Figure 7A, 7B) except invasion fitness for the AwsX mutant (two-
549 tailed t-tests $p < 0.08$, one outlier). This suggests that Pel polysaccharide serves as an
550 important structural component for colonizing the air-liquid interface and that its
551 production is activated by mutations leading to increased c-di-GMP levels. Although
552 deletion of *pel* in WS mutants resulted in less wrinkly colony morphology it did not
553 result in a smooth ancestral type. Neither did deletion of *pel* abolish the ability to
554 colonize the air liquid interface (Figure 8C) or the ability to invade wild type
555 populations (Figure 8A). This suggests that production of an additional EPS
556 component is induced by increased c-di-GMP levels caused by mutations in Wsp,
557 Aws and Mws, at least in the absence of *pel*. As expected if the motility defect
558 observed for WS mutants are primarily caused by high c-di-GMP levels rather than
559 high production of Pel, the motility was also reduced for Wsp, Aws and Mws mutants
560 with the *pel* operon deleted (Figure 8D).
561

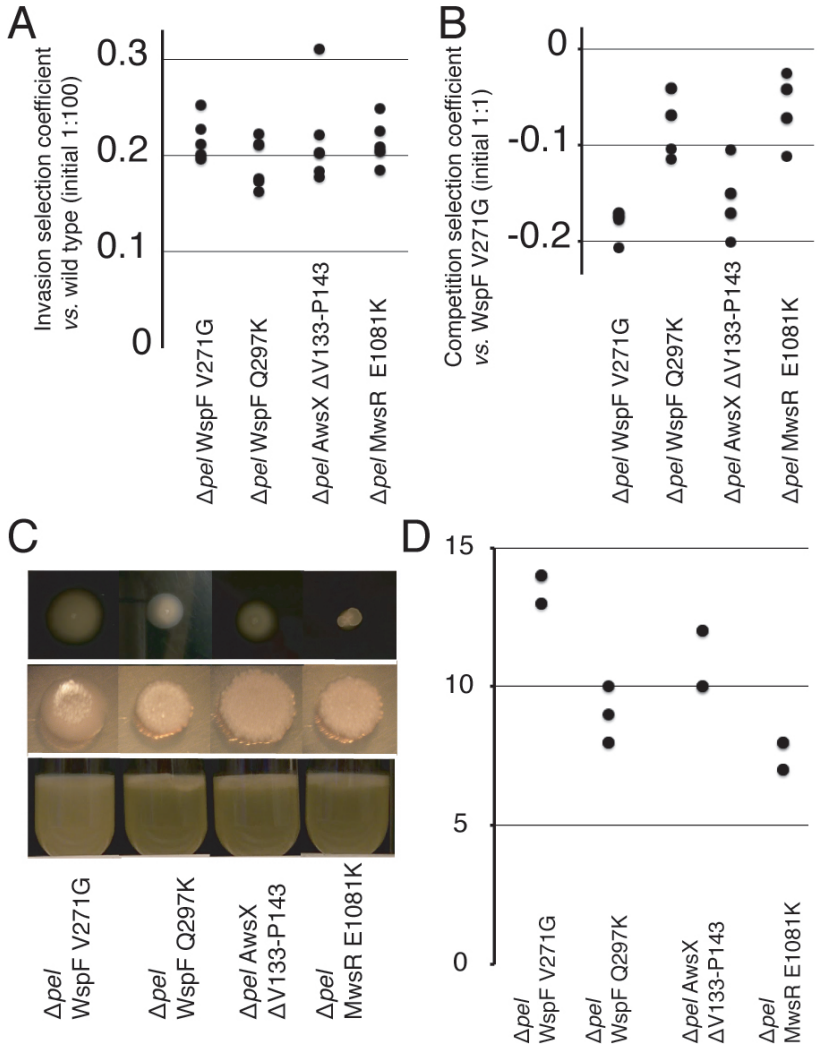


Fig 8. Contribution of *pel* to WS phenotype and fitness (**A**). Deletion of *pel* in WS mutants reduces invasion fitness. Fitness of reconstructed *P. protegens* Pf-5 WS mutants without the *pel* operon was measured in pairwise competitions. Invasion fitness was measured relative a dominant ancestral 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.

576 **Discussion**

577 The extension of the *P. fluorescens* SBW25 experimental evolution system to related
578 species shows promise for true testing of evolutionary forecasting models. While
579 there is a diversity of DGCs and EPSs between species leading to differences in
580 forecasts, the conserved role of c-di-GMP and limited number of phenotypes allow
581 the use of previous data to improve predictions and makes the experimental system
582 robust to changes in environmental conditions. The test of initial forecasts for *P.*
583 *protegens* Pf-5 presented here provides support for the ability to predict some aspects
584 of both genetic and phenotypic evolution while recognizing that the probability of
585 specific mutations cannot in most cases be predicted.
586
587 That experimental populations of *Pseudomonas* will colonize the air-liquid interface
588 when incubated under static condition is a prerequisite of extending the model. Given
589 that a range of phenotypic solutions is predicted to be available for all species the
590 evolution of such mutants for *P. protegens* is not surprising. The specific
591 environmental conditions used for experimental evolution often have a major impact
592 on evolutionary outcomes and is also likely to influence relative fitness and possibly
593 mutational biases also in the WS system. However despite major changes in growth
594 medium, temperature and material and physical dimensions of the growth vessel,
595 predictions on both the genetic and phenotypic levels proved successful
596 demonstrating robustness to environmental change and the establishment of a
597 dominant selective pressure, i.e. access to oxygen solved by air-liquid interface
598 colonization.
599
600 Phenotypic predictions of the structural basis supporting air-liquid colonization is
601 challenging given the limited previous experimental data. For SBW25 cellulose-based
602 solutions are superior in fitness, but for Pf-5 this solution in not available. The
603 prediction that overproduction of structural exopolysaccharides, rather than fuzzy,
604 cell-chaining or mucoid types, would be the primary solution was successful. One of
605 the two phenotypes found here used the Pel EPS, which could be predicted based on
606 its role in *P. aeruginosa*. However it appears to use a secondary EPS as well, that
607 remains to be identified, given that mutants lacking Pel but with activated DGCs still
608 colonize the air-liquid interface and have a distinct colony morphology. The second

609 phenotype used another EPS, encoded by PFL_3078-3093, which had not previously
610 be described and given that several EPS loci are usually encoded in *Pseudomonas*
611 genomes its use could not be predicted. However, repeating experimental evolution
612 using other *Pseudomonas* species is likely to provide more information about which
613 EPSs can be used to colonize the air-liquid interface and their relative fitness to allow
614 improved phenotypic predictions. Deletion of the *pel* operon, the unidentified
615 secondary EPS and PFL_3078-3093 and subsequent experimental evolution could
616 reveal less fit phenotypic solutions that are expected to exist including fuzzy types
617 caused by defects in LPS modification, cell-chaining types with defects in cell
618 division, adhesive proteins or mucoid types using alginate or levan, two EPSs with
619 lower structural stability.
620
621 The general prediction of types of mutations, as described in the hierarchy in Figure
622 2D, was also successful although the relatively few mutants identified here did not
623 allow for detection of rare activating mutations or double inactivating mutations. The
624 existence of such types could be confirmed by deletion of the main pathways (Wsp,
625 Aws, and Mws) followed by experimental evolution as previously described for
626 SBW25 (Lind, et al. 2015). The majority of mutations were loss-of-function
627 mutations in negative regulators or interacting proteins followed by less common
628 promoter mutations and promoter captures. In contrast to SBW25, where all promoter
629 mutations resulted in up-regulation of DGCs, the mutation upstream of PFL3078-
630 3093 demonstrates the possibility of direct transcriptional activation of EPS
631 components that are not under post-translational control of c-di-GMP. Two identical
632 mutations were found over 9 kb upstream of the *aws* operon, in between a ribosomal
633 RNA operon and the *recCBD* operon, which encodes key genes for recombination.
634 The molecular effects of these mutations have not been further investigated, but the
635 resulting WS phenotype is dependent of the presence of the *aws* operon, deletion of
636 which reversed the phenotype. This is consistent with an up-regulation of c-di-GMP
637 by AwsR presumably caused by increased transcription. The mutation is located in
638 the predicted terminator of the ribosomal RNA operon and increased transcriptional
639 read-through could put the *aws* operon under control of a very strong *rrn* promoter
640 that is most highly transcribed during exponential growth. This could explain the
641 relatively mild colony morphology phenotype as well as high motility of this WS
642 mutant (Figure 6).

643
644 The mathematical null model (Lind, et al. 2018) successfully predicted that Wsp
645 would be the most commonly used pathway followed by Aws, and Mws the most
646 rare. However, the number of mutants isolated here is rather small and the high
647 frequency of Wsp mutants seems mainly to be caused by a mutational hot spot in
648 *wspF*. Still the prediction that the three pathways together would contribute the large
649 majority of adaptive mutations (40 out of 43) is not trivial given that in SBW25 at
650 least 13 additional pathways are available to the high fitness WS phenotype (Lind, et
651 al. 2015). It is also worth noting that direct use of mutation rate data from SBW25
652 (Lind, et al. 2018) would result in poorer predictions than the mathematical null
653 model due to a strong mutational hot spot in *aws*X in that species.
654
655 For the multi-protein pathways Wsp and Aws, the null model predicted (Figure 2E)
656 that mutations would primarily be found in WspA, WspE, WspF, AwsX and AwsR.
657 Mutations were detected in all these except in WspA and the majority was found in
658 the negative regulators WspF and AwsX. WspA mutations were not found in the
659 original study in SBW25 either (McDonald, et al. 2009), but this was shown to be due
660 to lower fitness relative WspF and WspE mutants rather than a lower mutation rate to
661 WS (Lind, et al. 2018). Possibly this explains the absence of WspA mutants here as
662 well, but it is not clear if this fitness difference would be conserved in other species or
663 if sometimes WspA mutants are more fit. Thus the null model prediction of equal
664 rates for WspA, WspF and WspE is not changed for future experimental tests.
665
666 The molecular effects of the mutations found here are unknown, but knowledge from
667 SBW25 and *P. aeruginosa* and their positions in protein structure allowed some
668 predictions to be made. Inactivating mutations in the negative regulator WspF were
669 predicted to be either indels or missense mutations in four specific regions. Mutations
670 were found in two of the predicted regions, one in the vicinity to the methylesterase
671 active site where mutations are predicted to cause large disruptions in protein
672 structure and the other one directly disrupting the phosphorylation active site in the
673 signal receiver domain. No mutations were found in the surface exposed regions
674 hypothesized to be involved in interactions with WspA and WspE, which could be
675 due to differences in function between SBW25 and Pf-5 or simply that they appear at
676 lower frequency and would be detected if additional mutations were isolated. The sole

677 mutation in WspE is, as predicted, located in the direct vicinity of the phosphorylation
678 active site. Mutations in AwsX were amino acid substitutions throughout the gene as
679 well as in frame deletions inactivating the gene as predicted. Mutations in AwsR and
680 MwsR were also found in predicted regions, but no mutations were found in the
681 periplasmic region of AwsR, which is the most commonly targeted region in SBW25.
682 Known mutational hot spots in *awsX*, *awsR* and *mwsR* in SBW25 (Lind, et al. 2018)
683 were not conserved in Pf-5 resulting in divergent spectra of mutations, while mutated
684 regions and predicted functional effects remain conserved between the two species.
685
686 The diversity of phenotypic solutions observed after experimental evolution is
687 dependent on fitness differences between the phenotypes, but also on the rate of
688 which phenotypes are introduced by mutations, which is dependent on the genetic
689 architecture underlying the trait as well as mutational biases. The Pf-5 strain has at
690 least three DGC pathways (Wsp, Aws and Mws) that are subject to negative
691 regulation leading to prediction of a high rate of WS mutants, which are then expected
692 to outcompete other phenotypic solutions. If instead only one of these pathways were
693 present, a larger diversity of phenotypes would be expected to be observed with
694 relative fitness becoming less important as the first mutant that gains a foothold at the
695 air-liquid interface will have a large advantage and priority effects, i.e. being first,
696 will increasingly determine which adaptive mutants are observed.
697
698 Given that the mutational target upstream of PFL_3078-3093 is likely be relatively
699 small and that these mutants are rapidly outcompeted by all WS types tested, their
700 relatively high frequency (3/43) is unexpected. Possibly this is due to a higher
701 mutation rate at these sites (Sankar, et al. 2016) or that population structure limits
702 direct competition between these different phenotypes and reduces the importance of
703 relative fitness. In SBW25 low fitness phenotypes that colonize the air-liquid
704 interface based on LPS modification or cell-chaining are observed prior to the rise of
705 WS to high frequencies (Lind, et al. 2017b) due to the presence of mutational hot
706 spots in these genes which make these mutants appear early during the growth phase
707 despite their relatively small mutational targets (Ferguson, et al. 2013; Farr 2015;
708 Lind, et al. 2017b).
709

710 In successfully predicting evolutionary outcomes in *P. protegens*, this work lays the
711 foundation for future tests of evolutionary forecasting in related *Pseudomonas* species
712 by clearly stating predictions on several different levels from phenotype down to
713 which specific regions of proteins are likely to be mutated. Given what is already
714 known about the effects of (for now) unpredictable mutational biases and differences
715 in fitness between different WS types many of the forecasts will inevitably fail.
716 However hopefully they will fail in interesting ways thereby revealing erroneous
717 assumptions. The ability to remove common genetic and phenotypic pathways
718 provides a unique opportunity to also find those pathways that evolution does not
719 commonly use. This is necessary to determine why forecasts fail and update the
720 predictive models for the next cycle of prediction, experimental evolution and mutant
721 characterization to define the information necessary to predict short-term evolutionary
722 processes.
723

724 ## Materials and methods

725

726 **Strains and media**

727 *Pseudomonas protegens* Pf-5 (previously known as *P. fluorescens* Pf-5) and
728 derivatives thereof were used for all experimental evolution and phenotypic
729 characterization. *E. coli* DH5α was used for cloning PCR fragments for genetic
730 engineering (Paulsen et al). *P. protegens* Pf-5 was grown in tryptic soy broth
731 (Tryptone 17g, Soytone 3g, Glucose 2.5g, NaCl 5g, K_2HPO_4 2.5g per liter)
732 supplemented with 10 mM MgSO4 and 0.2% glycerol (TSBGM) for experimental
733 evolution and fitness assays. Lysogeny broth (LB) was used during genetic
734 engineering and LB without NaCl and supplemented with 8% sucrose was used for
735 counter-selection of *sacB* marker. Solid media were 1.5% agar added to LB or TSB
736 supplemented with 10 mM MgSO4, 0.2% glycerol and 10 mg/l Congo red. Motility
737 assays were conducted in 0.3% agar TSB supplemented with 10 mM MgSO4, 0.2%
738 glycerol. Kanamycin was used at 50 mg/l for *E. coli* or 80 mg/l for *P. protegens* and
739 gentamicin at 10 mg/l for *E. coli* or 15 mg/L for *P. protegens*. Selection plates for
740 cloning contained 5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside (X-gal) at 40
741 mg/l. 100 mg/L nitrofurantoin was used to inhibit growth of *E. coli* donor cells after
742 conjugation. All strains were stored at -80°C in LB with 10% DMSO.

743

744 **Experimental evolution**

745 30 central wells of a deep well plate (polypropylene, 1.1 mL, round walls, Axygen
746 Corning Life Sciences) were inoculated with approximately 10^3 cells each and
747 incubated at 36°C for 5 days without shaking on two different occasions. Suitable
748 dilutions were plated on TSBGM plates with Congo red after 5 days and incubated at
749 36°C for 48 h. Plates were screened for colonies with a visible difference in colony
750 morphology and one divergent colony per well were randomly selected based only on
751 its position on the agar plate. In total 43 independent mutants were streaked for single
752 cells twice before overnight growth in LB and freezing.

753

754 **Genome sequencing**

755 Seven mutant strains that did not contain mutations in the *wspF* and *awsX* genes were
756 analyzed by genome resequencing. The strains had mutations in *awsR*, *mwsR*, *wspE,*
757 upstream PFL_3078 (2 strains) and in the intergenic region between *rrfB* and *recC*
758 upstream of the awsXRO operon. Genomic DNA was isolated with Genomic DNA
759 Purification Kit (Thermo Fisher). Sequencing libraries were prepared from 1µg DNA
760 using the TruSeq PCRfree DNA sample preparation kit (cat# FC- 121-3001/3002,
761 Illumina Inc.) targeting an insert size of 350bp. The library preparation was
762 performed according to the manufacturers' instructions (guide#15036187).
763 Sequencing was performed with MiSeq (Illumina Inc.) paired-end 300bp read length
764 and v3 sequencing chemistry.

765

766 Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The
767 facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for
768 Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research
769 Council and the Knut and Alice Wallenberg Foundation. Sequencing data were
770 analyzed with using Geneious v. 10.2.3 with reads assembled against the *P. protegens*
771 Pf-5 genome sequence (CP000076.1).

772

773 **Sanger sequencing**

774 Sanger sequencing were performed by GATC biotech and used to sequence candidate
775 genes to find adaptive mutations and to confirm reconstructed mutations
776 (oligonucleotide primer sequences are available in Table S3).

777

778 **Reconstruction of mutations**

779 Nine mutations representing all candidate genes found using Sanger or Illumina
780 sequencing were reconstructed in the wild type ancestral *P. protegens* Pf-5 to show
781 that they are the cause of the adaptive phenotype and to be able to assay their fitness
782 effects without the risk of secondary mutations that might have occurred during
783 experimental evolution. A two-step allelic replacement protocol was using to transfer
784 the mutation into the ancestor. First a 1-2 kb fragment surrounding the putative
785 adaptive mutations were amplified using PCR (Phusion High- Fidelity DNA
786 polymerase, Thermo Scientific) and ligated into the multiple cloning site of the
787 mobilizable pK18mobsac suicide plasmid (FJ437239) using standard molecular
788 techniques. The ligation mix was then transformed into competent *E. coli* DH5α using
789 heat shock. After confirmation of correct insert size by PCR the plasmid was
790 transferred to *P. protegens* Pf-5 by conjugation with the donor strain and an *E. coli*
791 strain carrying the conjugation helper plasmid pRK2013. Cultures were grown
792 overnight of the recipient *P. protegens* Pf-5 (20 ml per conjugation at 30°C in LB)
793 and 2 ml each of the donor and helper *E. coli* strains per conjugation at 37°C in LB
794 with kanamycin. The culture of *P. protegens* Pf-5 was heat shocked for 10 minutes at
795 42°C prior to centrifugation at 4000 rpm for 10 minutes and resuspension in a small
796 volume of LB. Donor and helper cells were collected by centrifugation 4000 rpm for
797 10 minutes, resuspended in LB, and mixed with the concentrated recipient cells. After
798 another round of centrifugation the conjugation mix was resuspended in 50 µl LB and
799 spread onto several spots on a LA plate followed by incubation overnight at 30°C.
800 Each spot of the conjugation mix was scraped of the plate and resuspended in 200 µl
801 LB each and plated on LA plates with kanamycin, to select for transfer of the plasmid,
802 and nitrofurantoin that prevents growth of the *E. coli* donor and helper cells. The
803 pK18mobsac plasmid has a pBR322 type origin and cannot replicate in *P. protegens*
804 Pf-5 and only cells where the plasmid has integrated into the chromosome by
805 homologous recombination, with the homology provided by the cloned fragment, can
806 grow in the presence of kanamycin. After streaking for single cells on LA plates with
807 kanamycin, the *P. protegens* Pf-5 strains with integrated plasmids were grown
808 overnight in LB at 30°C without antibiotics to allow for double crossover
809 homologous recombination resulting in loss of the integrated plasmid. The plasmid

810 also contains the *sacB* marker conferring sucrose sensitivity, which allows for
811 counter-selection by plating on LA plates with sucrose. Sucrose resistant colonies
812 were checked for loss of the kanamycin marker and DNA sequencing of the cloned
813 region to find strains with the reconstructed mutation and no other mutations.
814 Deletion of the *pelABCDEFG* (PFL_2972-PFL_2978) region was accomplished using
815 the same two-step allelic exchange protocol using SOE-PCR to generated at fragment
816 surrounding the *pel* operon as previously described (Ferguson, et al. 2013; Farr 2015;
817 Lind, et al. 2017b). All oligonucleotide primer sequences are available in Table S3.
818
819 **Fitness assays**
820 Two types of competition fitness assays were performed similarly to previously
821 described (Ferguson, et al. 2013; Farr 2015; Lind, et al. 2015). The first assay
822 measures invasion fitness, where a mutant is mixed 1:100 with the wild type ancestor,
823 simulating early stages of air-liquid interface colonization where a rare mutant
824 establishes and grows at the surface with no competition from other mutants. The
825 second assay instead measures competition fitness in a 1:1 competition against a
826 reference mutant strain, which here was chosen to be the WspF V271G mutant
827 because it was the most commonly found in the experimental evolution study and thus
828 is highly successful either because of a high rate of emergence, i.e. a mutational hot
829 spot, or higher fitness than most other WS mutants. In addition, the WspF V271G
830 mutant has a temperature sensitive colony morphology phenotype in that it is highly
831 wrinkly at 30°C, but only have a very mild phenotype when grown at room
832 temperature, thus allowing it to be distinguishable from both the smooth ancestor and
833 all other wrinkly mutants isolated here.
834
835 Fluorescent reference strains of the wild type ancestor and the WspF V271G mutants
836 were created using a miniTn7 transposon (miniTn7(Gm) PA1/04/03 Gfp.AAV-a)
837 (Lambertsen, et al. 2004) that allows integration at a defined locus (attTn7) in the
838 chromosome. This allows the colonies to be distinguished not only by morphology
839 but also by fluorescence under blue/UV light and gentamicin resistance, which
840 provides a way to ascertain that secondary adaptive mutants that might occur during
841 the competition experiment do not bias the results (for example the ancestor could
842 evolve WS types or a WS mutant can evolve to cheat on the other type by inactivation
843 of EPS production or reduced c-di-GMP signalling). Introduction of the transposon

844 into *P. protegens* Pf-5 was performed by tri-parental conjugation from *E. coli* with
845 helper plasmids pRK2013 (conjugation helper) and pUX-BF13 containing the
846 transposase genes) using the same conjugation protocol described above.
847
848 The invasion assay was performed by mixing shaken overnight cultures of the
849 competitor 1:100 with the GFP-labeled reference ancestor followed by 1000-fold
850 dilution and static incubation at 36°C for 48 h in TBSGM medium in deep well plates
851 (1 ml per well, using only the central 60 wells). For the competition assay, the GFP-
852 labeled reference strain WspF 271G was mixed 1:1 with the competitor and diluted 6-
853 fold and grown for 4 h (shaken at 30°C), before plating to determine initial ratios, to
854 ensure the cells were in a similar physiological state at the start of the competition.
855 The competition cultures were then diluted 1000-fold in TBSGM medium and grown
856 in deep well plates (1 ml per well, using only the central 60 wells) static for 24 h at
857 36°C. Selection coefficients (s) were calculated as previously described (Dykhuizen
858 1990), where s = 0 is equal fitness, positive is increased fitness and negative is
859 decreased fitness relative to the reference strain. Briefly s is calculated as the change
860 in logarithmic ratio over time according to $s = [\ln(R(t)/R(0))]/[t]$, where R is the ratio
861 of mutant to reference and t is the number of generations of the entire population
862 during the experiment (estimated from viable counts). The cost of the fluorescent
863 marker were calculated from control competitions where the GFP-labeled reference
864 strains (wild type and WspF V271G) were competed against isogenic strains without
865 the marker and included in each plate under identical conditions during the fitness
866 assays and used to adjust the selection coefficients to compensate for the cost.
867
868 **Motility assays**
869 Swimming motility assays were performed in TBSGM plates with 0.3% agar (BD)
870 and the diameter was measured after 24 h of growth at room temperature. Each strain
871 was assayed in duplicates on two different plates.
872
873 **Bioinformatics analysis of DGCs and EPS genes**
874 Homologs for all DGCs in *P. fluorescens* SBW25 were found using the *Pseudomonas*
875 Ortholog Database at Pseudomonas.com (Winsor, et al. 2016). Blast-p searches for
876 GGDEF domains were performed to find remaining DGCs in the six *Pseudomonas*

877 species and their homologs again found using the *Pseudomonas* Ortholog Database
878 (Whiteside, et al. 2013) and manually inspected. Annotations (Pseudomonas.com. DB
879 version 17.2) were also searched for diguanylate cyclase and GGDEF. Not all DCCs
880 found are likely to have diguanylate cyclase activity, but given the difficulties of
881 predicting which of the partly degenerate active sites are likely to be inactive
882 combined with the possibilities of mutational activation during experimental
883 evolution, none were excluded.
884
885 There is no simple way to find all genes that can function as structural or regulatory
886 genes to allow colonization of the air-liquid interface. Thus the selection in Figure 3B
887 and Figure 3 – source data should not be considered complete. Putative EPS genes
888 were found using blastp searches with sequences from known exopolysaccharide
889 biosynthesis proteins including cellulose, PGA, Pel, Psl, Pea, Peb, alginate and levan.
890 Homologs were then found using the *Pseudomonas* Ortholog Database (Whiteside, et
891 al. 2013) at Pseudomonas.com (Winsor, et al. 2016). Annotations (Pseudomonas.com.
892 DB version 17.2) were also searched for glycosyltransferase, glycosyl transferase,
893 flippase, polysaccharide, lipopolysaccharide, polymerase, biofilm, adhesion and
894 adhesion. Based on previous work in SBW25 and literature searches a few additional
895 genes were added.
896

897 ## Acknowledgements

898 This work was supported by grants from Carl Tryggers Foundation for Scientific
899 Research and Magnus Bergvalls Foundation.
900

901 ## Competing interests

902 The author declares no competing interests.

903 ## References

904

905 Bantinaki E, Kassen R, Knight CG, Robinson Z, Spiers AJ, Rainey PB. 2007.
906 Adaptive divergence in experimental populations of Pseudomonas fluorescens. III.
907 Mutational origins of wrinkly spreader diversity. Genetics 176:441-453.
908 http://dx.doi.org/10.1534/genetics.106.069906

909 Barrick JE, Lenski RE. 2013. Genome dynamics during experimental evolution.
910 Nature reviews. Genetics 14:827-839. http://dx.doi.org/10.1038/nrg3564

911 Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF.
912 2009. Genome evolution and adaptation in a long-term experiment with
913 Escherichia coli. Nature 461:1243-1247. http://dx.doi.org/10.1038/nature08480

914 Beaumont HJ, Gallie J, Kost C, Ferguson GC, Rainey PB. 2009. Experimental
915 evolution of bet hedging. Nature 462:90-93. http://dx.doi.org/10.1038/nature08504

916 Blank D, Wolf L, Ackermann M, Silander OK. 2014. The predictability of molecular
917 evolution during functional innovation. Proceedings of the National Academy of
918 Sciences of the United States of America 111:3044-3049.
919 http://dx.doi.org/10.1073/pnas.1318797111

920 Brandis G, Pietsch F, Alemayehu R, Hughes D. 2015. Comprehensive phenotypic
921 characterization of rifampicin resistance mutations in Salmonella provides insight
922 into the evolution of resistance in Mycobacterium tuberculosis. J Antimicrob
923 Chemother 70:680-685. http://dx.doi.org/10.1093/jac/dku434

924 Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, Palsson BO. 2009.
925 Whole-genome resequencing of Escherichia coli K-12 MG1655 undergoing short-
926 term laboratory evolution in lactate minimal media reveals flexible selection of
927 adaptive mutations. Genome biology 10:R118. http://dx.doi.org/10.1186/gb-2009-
928 10-10-r118

929 Deatherage DE, Kepner JL, Bennett AF, Lenski RE, Barrick JE. 2017. Specificity of
930 genome evolution in experimental populations of Escherichia coli evolved at
931 different temperatures. Proceedings of the National Academy of Sciences of the
932 United States of America 114:E1904-E1912.
933 http://dx.doi.org/10.1073/pnas.1616132114

934 Dykhuizen DE. 1990. Experimental studies of natural selection in bacteria. Annu.
935 Rev. Ecol. Syst. 21:373-398.

936 Farr AD. 2015. Formalist features determining the tempo and mode of evolution in
937 Pseudomonas fluorescens SBW25. [[Auckland, New Zealand]: Massey University.

938 Farr AD, Remigi P, Rainey PB. 2017. Adaptive evolution by spontaneous domain
939 fusion and protein relocalization. Nat Ecol Evol 1:1562-1568.
940 http://dx.doi.org/10.1038/s41559-017-0283-7

941 Ferguson GC, Bertels F, Rainey PB. 2013. Adaptive Divergence in Experimental
942 Populations of Pseudomonas fluorescens. V. Insight into the Niche Specialist
943 "Fuzzy Spreader" Compels Revision of the Model Pseudomonas Radiation.
944 Genetics. http://dx.doi.org/10.1534/genetics.113.154948

945 Firnberg E, Labonte JW, Gray JJ, Ostermeier M. 2014. A comprehensive, high-
946 resolution map of a gene's fitness landscape. Molecular biology and evolution
947 31:1581-1592. http://dx.doi.org/10.1093/molbev/msu081

948 Fraebel DT, Mickalide H, Schnitkey D, Merritt J, Kuhlman TE, Kuehn S. 2017.
949 Environment determines evolutionary trajectory in a constrained phenotypic space.
950 Elife 6. http://dx.doi.org/10.7554/eLife.24669

951 Friedman L, Kolter R. 2004. Genes involved in matrix formation in Pseudomonas
952 aeruginosa PA14 biofilms. Mol Microbiol 51:675-690.

953 Gallie J, Libby E, Bertels F, Remigi P, Jendresen CB, Ferguson GC, Desprat N,
954 Buffing MF, Sauer U, Beaumont HJ, et al. 2015. Bistability in a metabolic network
955 underpins the de novo evolution of colony switching in Pseudomonas fluorescens.
956 PLoS biology 13:e1002109. http://dx.doi.org/10.1371/journal.pbio.1002109

957 Goymer P, Kahn SG, Malone JG, Gehrig SM, Spiers AJ, Rainey PB. 2006. Adaptive
958 divergence in experimental populations of Pseudomonas fluorescens. II. Role of
959 the GGDEF regulator WspR in evolution and development of the wrinkly spreader
960 phenotype. Genetics 173:515-526. http://dx.doi.org/10.1534/genetics.106.055863

961 Herron MD, Doebeli M. 2013. Parallel evolutionary dynamics of adaptive
962 diversification in Escherichia coli. PLoS biology 11:e1001490.
963 http://dx.doi.org/10.1371/journal.pbio.1001490

964 Hickman JW, Tifrea DF, Harwood CS. 2005. A chemosensory system that regulates
965 biofilm formation through modulation of cyclic diguanylate levels. Proceedings of
966 the National Academy of Sciences of the United States of America 102:14422-
967 14427. http://dx.doi.org/10.1073/pnas.0507170102

968 Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, Bercot B, Petit
969 E, Poulain J, Barnaud G, et al. 2013. Capturing the mutational landscape of the
970 beta-lactamase TEM-1. Proceedings of the National Academy of Sciences of the
971 United States of America 110:13067-13072.
972 http://dx.doi.org/10.1073/pnas.1215206110

973 Jahn LJ, Munck C, Ellabaan MMH, Sommer MOA. 2017. Adaptive Laboratory
974 Evolution of Antibiotic Resistance Using Different Selection Regimes Lead to
975 Similar Phenotypes and Genotypes. Front Microbiol 8:816.
976 http://dx.doi.org/10.3389/fmicb.2017.00816

977 Jerison ER, Desai MM. 2015. Genomic investigations of evolutionary dynamics and
978 epistasis in microbial evolution experiments. Curr Opin Genet Dev 35:33-39.
979 http://dx.doi.org/10.1016/j.gde.2015.08.008

980 Jimenez JI, Xulvi-Brunet R, Campbell GW, Turk-MacLeod R, Chen IA. 2013.
981 Comprehensive experimental fitness landscape and evolutionary network for small
982 RNA. Proceedings of the National Academy of Sciences of the United States of
983 America 110:14984-14989. http://dx.doi.org/10.1073/pnas.1307604110

984 Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web
985 portal for protein modeling, prediction and analysis. Nat Protoc 10:845-858.
986 http://dx.doi.org/10.1038/nprot.2015.053

987 Kishimoto T, Iijima L, Tatsumi M, Ono N, Oyake A, Hashimoto T, Matsuo M,
988 Okubo M, Suzuki S, Mori K, et al. 2010. Transition from positive to neutral in
989 mutation fixation along with continuing rising fitness in thermal adaptive
990 evolution. PLoS genetics 6:e1001164.
991 http://dx.doi.org/10.1371/journal.pgen.1001164

992 Knoppel A, Knopp M, Albrecht LM, Lundin E, Lustig U, Nasvall J, Andersson DI.
993 2018. Genetic Adaptation to Growth Under Laboratory Conditions in Escherichia
994 coli and Salmonella enterica. Front Microbiol 9:756.
995 http://dx.doi.org/10.3389/fmicb.2018.00756

996 Koskiniemi S, Sun S, Berg OG, Andersson DI. 2012. Selection-driven gene loss in
997 bacteria. PLoS genetics 8:e1002787.
998 http://dx.doi.org/10.1371/journal.pgen.1002787

999 Kram KE, Geiger C, Ismail WM, Lee H, Tang H, Foster PL, Finkel SE. 2017.
1000 Adaptation of Escherichia coli to Long-Term Serial Passage in Complex Medium:
1001 Evidence of Parallel Evolution. mSystems 2.
1002 http://dx.doi.org/10.1128/mSystems.00192-16

1003 LaCroix RA, Sandberg TE, O'Brien EJ, Utrilla J, Ebrahim A, Guzman GI, Szubin R,
1004 Palsson BO, Feist AM. 2015. Use of adaptive laboratory evolution to discover key
1005 mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose
1006 minimal medium. Appl Environ Microbiol 81:17-30.
1007 http://dx.doi.org/10.1128/AEM.02246-14

1008 Lambertsen L, Sternberg C, Molin S. 2004. Mini-Tn7 transposons for site-specific
1009 tagging of bacteria with fluorescent proteins. Environmental microbiology 6:726-
1010 732. http://dx.doi.org/10.1111/j.1462-2920.2004.00605.x

1011 Lee MC, Marx CJ. 2012. Repeated, selection-driven genome reduction of accessory
1012 genes in experimental populations. PLoS genetics 8:e1002651.
1013 http://dx.doi.org/10.1371/journal.pgen.1002651

1014 Lee VT, Matewish JM, Kessler JL, Hyodo M, Hayakawa Y, Lory S. 2007. A cyclic-
1015 di-GMP receptor required for bacterial exopolysaccharide production. Mol
1016 Microbiol 65:1474-1484. http://dx.doi.org/10.1111/j.1365-2958.2007.05879.x

1017 Liang ZX. 2015. The expanding roles of c-di-GMP in the biosynthesis of
1018 exopolysaccharides and secondary metabolites. Nat Prod Rep 32:663-683.
1019 http://dx.doi.org/10.1039/c4np00086b

1020 Lind PA, Arvidsson L, Berg OG, Andersson DI. 2017a. Variation in Mutational
1021 Robustness between Different Proteins and the Predictability of Fitness Effects.
1022 Molecular biology and evolution 34:408-418.
1023 http://dx.doi.org/10.1093/molbev/msw239

1024 Lind PA, Farr AD, Rainey PB. 2017b. Evolutionary convergence in experimental
1025 Pseudomonas populations. ISME J 11:589-600.
1026 http://dx.doi.org/10.1038/ismej.2016.157

1027 Lind PA, Farr AD, Rainey PB. 2015. Experimental evolution reveals hidden diversity
1028 in evolutionary pathways. Elife 4. http://dx.doi.org/10.7554/eLife.07074

1029 Lind PA, Libby E, Herzog J, Rainey PB. 2018. Disentangling the effects of genetic
1030 architecture, mutational bias and selection on evolutionary forecasting. bioRxiv.
1031 http://dx.doi.org/10.1101/335711

1032 Long A, Liti G, Luptak A, Tenaillon O. 2015. Elucidating the molecular architecture
1033 of adaptation via evolve and resequence experiments. Nature reviews. Genetics
1034 16:567-582. http://dx.doi.org/10.1038/nrg3937

1035 Lukacisinova M, Bollenbach T. 2017. Toward a quantitative understanding of
1036 antibiotic resistance evolution. Curr Opin Biotechnol 46:90-97.
1037 http://dx.doi.org/10.1016/j.copbio.2017.02.013

1038 Lundin E, Tang PC, Guy L, Nasvall J, Andersson DI. 2017. Experimental
1039 determination and prediction of the fitness effects of random point mutations in the
1040 biosynthetic enzyme HisA. Molecular biology and evolution.
1041 http://dx.doi.org/10.1093/molbev/msx325

1042 McDonald MJ, Cooper TF, Beaumont HJE, Rainey PB. 2011. The distribution of
1043 fitness effects of new beneficial mutations in Pseudomonas fluorescens. Biology
1044 Letters 7:98-100. http://dx.doi.org/Doi 10.1098/Rsbl.2010.0547

1045 McDonald MJ, Gehrig SM, Meintjes PL, Zhang XX, Rainey PB. 2009. Adaptive
1046 divergence in experimental populations of Pseudomonas fluorescens. IV. Genetic
1047 constraints guide evolutionary trajectories in a parallel adaptive radiation. Genetics
1048 183:1041-1053. http://dx.doi.org/10.1534/genetics.109.107110

1049 McElroy KE, Hui JG, Woo JK, Luk AW, Webb JS, Kjelleberg S, Rice SA, Thomas
1050 T. 2014. Strain-specific parallel evolution drives short-term diversification during
1051 Pseudomonas aeruginosa biofilm formation. Proceedings of the National Academy
1052 of Sciences of the United States of America 111:E1419-1427.
1053 http://dx.doi.org/10.1073/pnas.1314340111

1054 Morgan JL, McNamara JT, Zimmer J. 2014. Mechanism of activation of bacterial
1055 cellulose synthase by cyclic di-GMP. Nat Struct Mol Biol 21:489-496.
1056 http://dx.doi.org/10.1038/nsmb.2803

1057 O'Neill AJ, Huovinen T, Fishwick CW, Chopra I. 2006. Molecular genetic and
1058 structural modeling studies of Staphylococcus aureus RNA polymerase and the

1059 fitness of rifampin resistance genotypes in relation to clinical prevalence.
1060 Antimicrob Agents Chemother 50:298-309.
1061 http://dx.doi.org/10.1128/AAC.50.1.298-309.2006

1062 Orgogozo V. 2015. Replaying the tape of life in the twenty-first century. Interface
1063 Focus 5:20150057. http://dx.doi.org/10.1098/rsfs.2015.0057

1064 Rainey PB, Travisano M. 1998. Adaptive radiation in a heterogeneous environment.
1065 Nature 394:69-72. http://dx.doi.org/10.1038/27900

1066 Romling U, Galperin MY, Gomelsky M. 2013. Cyclic di-GMP: the first 25 years of a
1067 universal bacterial second messenger. Microbiology and molecular biology
1068 reviews : MMBR 77:1-52. http://dx.doi.org/10.1128/MMBR.00043-12

1069 Sandberg TE, Pedersen M, LaCroix RA, Ebrahim A, Bonde M, Herrgard MJ, Palsson
1070 BO, Sommer M, Feist AM. 2014. Evolution of Escherichia coli to 42 degrees C
1071 and subsequent genetic engineering reveals adaptive mechanisms and novel
1072 mutations. Molecular biology and evolution 31:2647-2662.
1073 http://dx.doi.org/10.1093/molbev/msu209

1074 Sankar TS, Wastuwidyaningtyas BD, Dong Y, Lewis SA, Wang JD. 2016. The nature
1075 of mutations induced by replication-transcription collisions. Nature.
1076 http://dx.doi.org/10.1038/nature18316

1077 Schenk MF, Szendro IG, Krug J, de Visser JA. 2012. Quantifying the adaptive
1078 potential of an antibiotic resistance enzyme. PLoS genetics 8:e1002783.
1079 http://dx.doi.org/10.1371/journal.pgen.1002783

1080 Silby MW, Cerdeno-Tarraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston
1081 GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SA, et al. 2009. Genomic and
1082 genetic analyses of diversity and plant interactions of Pseudomonas fluorescens.
1083 Genome biology 10:R51. http://dx.doi.org/10.1186/gb-2009-10-5-r51

1084 Sleight SC, Orlic C, Schneider D, Lenski RE. 2008. Genetic basis of evolutionary
1085 adaptation by Escherichia coli to stressful cycles of freezing, thawing and growth.
1086 Genetics 180:431-443. http://dx.doi.org/10.1534/genetics.108.091330

1087 Sommer MOA, Munck C, Toft-Kehler RV, Andersson DI. 2017. Prediction of
1088 antibiotic resistance: time for a new preclinical paradigm? Nature reviews.
1089 Microbiology 15:689-696. http://dx.doi.org/10.1038/nrmicro.2017.75

1090 Spiers AJ, Bohannon J, Gehrig SM, Rainey PB. 2003. Biofilm formation at the air-
1091 liquid interface by the Pseudomonas fluorescens SBW25 wrinkly spreader requires
1092 an acetylated form of cellulose. Molecular Microbiology 50:15-27.

1093 Spiers AJ, Kahn SG, Bohannon J, Travisano M, Rainey PB. 2002. Adaptive
1094 divergence in experimental populations of Pseudomonas fluorescens. I. Genetic
1095 and phenotypic bases of wrinkly spreader fitness. Genetics 161:33-46.

1096 Spiers AJ, Rainey PB. 2005. The Pseudomonas fluorescens SBW25 wrinkly spreader
1097 biofilm requires attachment factor, cellulose fibre and LPS interactions to maintain
1098 strength and integrity. Microbiology 151:2829-2839.
1099 http://dx.doi.org/10.1099/mic.0.27984-0

1100 Steiner S, Lori C, Boehm A, Jenal U. 2013. Allosteric activation of exopolysaccharide
1101 synthesis through cyclic di-GMP-stimulated protein-protein interaction. EMBO J
1102 32:354-368. http://dx.doi.org/10.1038/emboj.2012.315

1103 Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD,
1104 Gaut BS. 2012. The molecular diversity of adaptive convergence. Science
1105 335:457-461. http://dx.doi.org/10.1126/science.1212986

1106 Whiteside MD, Winsor GL, Laird MR, Brinkman FS. 2013. OrtholugeDB: a bacterial
1107 and archaeal orthology resource for improved comparative genomic analysis.
1108 Nucleic Acids Res 41:D366-376. http://dx.doi.org/10.1093/nar/gks1241

1109 Whitney JC, Whitfield GB, Marmont LS, Yip P, Neculai AM, Lobsanov YD,
1110 Robinson H, Ohman DE, Howell PL. 2015. Dimeric c-di-GMP is required for
1111 post-translational regulation of alginate production in Pseudomonas aeruginosa. J
1112 Biol Chem 290:12451-12462. http://dx.doi.org/10.1074/jbc.M115.645051

1113 Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ. 1999. Different
1114 trajectories of parallel evolution during viral adaptation. Science 285:422-424.

1115 Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FS. 2016. Enhanced
1116 annotations and features for comparing thousands of Pseudomonas genomes in the

1117 Pseudomonas genome database. Nucleic Acids Res 44:D646-653.
1118 http://dx.doi.org/10.1093/nar/gkv1227

1119

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

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			

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				

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

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

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

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
***aws* operon**	
RecC_F_HindIII	tagaagcttctgataccgcccaagagttc
RecC_R_KpnI	catggtaccccagtcggctcgatatacct
recD_F_EcoRI	acggaattcccactacctgaatgtactg
PFL_0743-SalI_F	tgtgtgctgtcgaccattc
awsR_182R	agctgatggagcgggcgatca
awsR_677F	cgacttcaacgccctgct
awsO_DR	ttacatccgcgaggtgac
PFL_0746_F_HindIII	tgcaagcttcttcttcatcaccccgcaa
mwsR	
mwsR_HindIII_F	gttaagcttagacccagctgttcctgttc
mwsR_2144F	gccgcgacatcagccagca
PFL_5345_SeqF	gaaaaggacctgcgcatg
PFL_5345_SeqR	gaactgcttgaggtagttc
mwsR_3599R	gaaggtgcggtcgatcttca
mwsR_SalI_R	aggccgtcgacgaaggtgc
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