Predicting mutational routes to new adaptive phenotypes

Peter A. Lind [1,2], Eric Libby [1,3,4], Jenny Herzog[1] and Paul B. Rainey[1,5,6]

[1]New Zealand Institute for Advanced Study, Massey University at Albany, Auckland, 0745, New Zealand.
[2]Department of Molecular Biology, Umeå University, SE-901 87 Umeå, Sweden.
[3]Santa Fe Institute, Santa Fe, New Mexico, United States of America.
[4]Department of Mathematics, Umeå University, SE-901 87 Umeå, Sweden.
[5]Department of Microbial Population Biology, Max Planck Institute for Evolutionary Biology, Plön 24306, Germany.
[6]Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI Paris-Tech), CNRS UMR 8231, PSL Research University, 75231 Paris, France.

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 fluorescens, experimental evolution, genetic architecture, wrinkly spreader, c-di-GMP, evolutionary forecasting

Short running title: complexities of the genotype-to-phenotype map

Impact statement: A combination of genetics, experimental evolution and mathematical modelling defines information necessary to predict the outcome of short-term adaptive evolution.

Abstract

Predicting evolutionary change poses numerous challenges. Here we take advantage of the model bacterium *Pseudomonas fluorescens* in which the genotype-to-phenotype map determining evolution of the adaptive "wrinkly spreader" (WS) type is known. We present mathematical descriptions of three necessary regulatory pathways and use these to predict both the rate at which each mutational route is used and the expected mutational targets. To test predictions, mutation rates and targets were determined for each pathway. Unanticipated mutational hotspots caused experimental observations to depart from predictions but additional data led to refined models. A mismatch was observed between the spectra of WS-causing mutations obtained with and without selection due to low fitness of previously undetected WS-causing mutations. Our findings contribute toward the development of mechanistic models for forecasting evolution, highlight current limitations, and draw attention to challenges in predicting locus-specific mutational biases and fitness effects.

Introduction

Adaptation requires the realization of beneficial mutations. As self-evident as this may be, predicting the occurrence of beneficial mutations and their trajectories to improved fitness is fraught with challenges (Lässig, et al. 2017). Nonetheless progress has been made for phenotypically diverse asexual populations subject to strong selection. Effective approaches have drawn upon densely sampled sequence data and equilibrium models of molecular evolution to predict amino acid preferences at specific loci (Łuksza and Lässig 2014). Predictive strategies have also been developed based on selection inferred from the shape of coalescent trees (Neher, et al. 2014). In both instances the models are coarse-grained and sidestep specific molecular and mutational details.

There is reason to by-pass molecular details: mutation, being a stochastic process, means that for the most part details are likely to be idiosyncratic and unpredictable. But an increasing number of investigations give reason to think otherwise – that adaptive molecular evolution might follow rules (Yampolsky and Stoltzfus 2001; Pigliucci 2010; Stern 2013; Laland, et al. 2015). This is particularly apparent from studies of parallel molecular evolution (Colosimo, et al. 2005; Woods, et al. 2006; Ostrowski, et al. 2008; Flowers, et al. 2009; Meyer, et al. 2012; Tenaillon, et al. 2012; Zhen, et al. 2012; Herron and Doebeli 2013; Galen, et al. 2015; Bailey, et al. 2017; Kram, et al. 2017; Stoltzfus and McCandlish 2017), and particularly from studies that show parallel evolution to be attributable – at least in part – to factors other than selection (McDonald, et al. 2009; Lind, et al. 2015; Bailey, et al. 2017).

A standard starting position for predicting adaptive evolution recognises the

importance of population genetic parameters including mutation rate, generation time, population size, selection and more recently information on the distribution of beneficial fitness effects. An often used approach appeals to "origin-fixation" models that estimate the probability that selection will realise particular mutational trajectories based on the assumption that the rate of evolution is a function of the rate at which mutations originate, multiplied by their fixation probability (for review, see (McCandlish and Stoltzfus 2014), for application, see for example (Weinreich, et al. 2006)). Leaving aside a number of simplifying and restrictive assumptions, population genetic approaches such as those encompassed by origin-fixation models require pre-existing knowledge of (or assumptions concerning) mutation rate to a particular phenotype and selection coefficients (obtained in the absence of frequency dependent effects) and thus have predictive power only in specific situations and with the benefit of hindsight (McCandlish and Stoltzfus 2014).

Looking to the future, there is hope that one day it might be possible to predict the course of evolutionary change in response to specific environmental challenges in much the same way as gene function, regulation and interactions can be predicted today based on knowledge of nucleotide sequence data. A central issue is to define the necessary information. Mechanistic understanding of the connection between genotype and phenotype combined with knowledge of the genomic bases of mutational bias offer opportunities for progress.

Mutations arise randomly with respect to utility, but genetic architecture can influence the translation of mutation into phenotypic variation: the likelihood that a given mutation generates phenotypic effects depends on the genotype-to-phenotype map

95 (Alberch 1991; Gompel and Prud'homme 2009; Stern and Orgogozo 2009; Rainey, et
96 al. 2017). Thus, the functions of gene products and their regulatory interactions
97 provide information on likely mutational targets underpinning particular phenotypes.
98 For example, consider a hypothetical structural determinant subject to both positive
99 and negative regulation and whose over-expression generates a given adaptive
100 phenotype. Assuming a uniform distribution of mutational events, mutations in the
101 negative regulator (and not the positive activator) will be the primary cause of the
102 adaptive phenotype. This follows from the fact that loss-of-function mutations are
103 more common than gain-of-function mutations. Indeed, an emerging rule indicates
104 that phenotypes determined by genetic pathways that are themselves subject to
105 negative regulation are most likely to arise by loss-of-function mutations in negative
106 regulatory components (McDonald, et al. 2009; Tenaillon, et al. 2012; Lind, et al.
107 2015; Fraebel, et al. 2017).
108
109 Mutation is not equally likely at each nucleotide of a given genome (Lind and
110 Andersson 2008; Lynch 2010; Seier, et al. 2011; Foster, et al. 2015; Reijns, et al.
111 2015; Sankar, et al. 2016; Stoltzfus and McCandlish 2017). Numerous instances of
112 mutational bias have been reported. Prime examples are simple sequence repeats
113 such as homopolymeric nucleotide tracts or di-, tri- and tetrameric repeats that mutate
114 at high frequency via slipped strand mispairing (Levinson and Gutman 1987). These
115 readily identifiable sequences define contingency loci in obligate human pathogens
116 and commensals (Moxon, et al. 1994) and are widespread in eukaryotic genomes
117 (Tautz and Renz 1984). The behaviour of contingency loci can be further modulated
118 by defects in components of methyl-directed mismatch repair systems (Richardson
119 and Stojiljkovic 2001; Martin, et al. 2004; Hammerschmidt, et al. 2014; Heilbron, et

120 al. 2014).

121

122 Certain palindromic structures also lead to mutational bias (Viswanathan, et al. 2000;
123 Lovett 2004) and promote amplification events that can increase target size for
124 additional mutations (Roth, et al. 1996; Kugelberg, et al. 2010; Reams and Roth 2015)
125 Transition-transversion bias (Stoltzfus and McCandlish 2017) and elevated mutation
126 rates at CpG sites (Galen, et al. 2015) can further skew the distributions of mutational
127 effects. Additional bias arises from the chromosomal neighbourhood of genes under
128 selection (Steinrueck and Guet 2017), the location of genes with regard to interactions
129 with DNA replication/transcription machineries (Sankar, et al. 2016), and
130 environmental factors that affect not only mutation rate but also the spectra of
131 mutational events (Krasovec, et al. 2017; Maharjan and Ferenci 2017; Shewaramani,
132 et al. 2017).

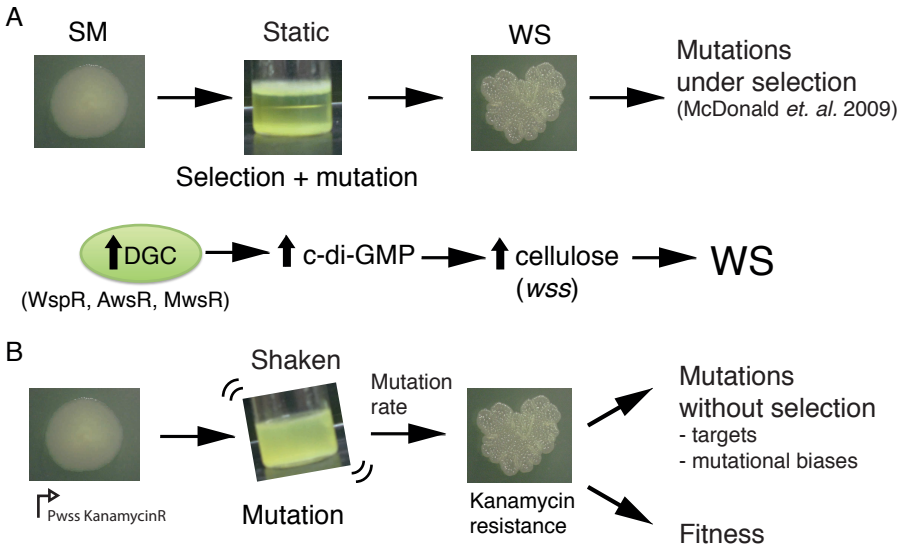
133

134 Beyond the genotype-to-phenotype map and mutational biases, predicting adaptive
135 evolution requires ability to know *a priori* the fitness effects of specific mutations. At
136 the present time there is much theoretical and empirical interest in the distribution of
137 fitness effects (DFE) (Eyre-Walker and Keightley 2007) — and particularly the DFE
138 of beneficial mutations (Orr 2005) — because of implications for predicting the rate
139 of adaption and likelihood of parallel evolution (de Visser and Krug 2014), but
140 knowledge of the shape of the distribution is insufficient to connect specific mutations
141 to their specific effects, or to their likelihood of occurrence. Such connections require
142 a means of knowing the connection between mutations and their environment-specific
143 fitness effects. This is a tall order. A starting point is to understand the relationship
144 between all possible mutational routes to a particular phenotype and the set that are

145 realised by selection.

146

147 Here we take a bacterial system in which the genetic pathways underpinning
148 evolution of the adaptive "wrinkly spreader" (WS) type are known and use this to
149 explore the current limits on evolutionary forecasting. *Pseudomonas fluorescens*
150 SBW25 growing in static broth microcosms rapidly depletes available oxygen
151 establishing selective conditions that favour mutants able to colonise the air-liquid
152 interface. The most successful mutant-class encompasses the WS types (Ferguson, et
153 al. 2013; Lind, et al. 2017b). These types arise from mutational activation of
154 diguanylate cyclases (DGCs) that cause over-production of the second messenger c-
155 di-GMP (Goymer, et al. 2006; McDonald, et al. 2009), over-production of an
156 acetylated cellulose polymer (Spiers, et al. 2002; Spiers, et al. 2003) and ultimately
157 formation of a self-supporting microbial mat (Figure 1A).



158

159 **Figure 1. The *Pseudomonas fluorescens* SBW25 wrinkly spreader model.** (**A**)
160 Selection for access to oxygen allows wrinkly spreader (WS) mutants to invade the
161 ancestral smooth (SM) population in static microcosms. WS mutants form a mat at
162 the air-liquid interface through increased expression of the main structural

163 component, cellulose, encoded by the *wss* operon. Expression of cellulose is
164 controlled by the second messenger c-di-GMP, which is produced by diguanylate
165 cyclases (DGCs). Mutations in the *wsp*, *aws* and *mws* operons, that activate their
166 respective DGCs (WspR, AwsR, MwsR), are the primary mutational pathways to WS.
167 (**B**) When a reporter construct connecting expression of the *wss* operon to resistance
168 to kanamycin is used under shaken non-selective conditions, WS mutants can be
169 isolated without the biasing influence of natural selection. This allows estimation of
170 the mutation rate to WS and an unbiased spectrum of mutations defining the
171 mutational target. Fitness can then be assayed in competition with a common
172 reference strain.
173
174 McDonald *et al.* (McDonald, et al. 2009) showed that each time the tape of WS
175 evolution is re-run mutations generating the adaptive type arise in one of three DGC-
176 encoding pathways (Wsp, Aws, or Mws) (Figure 1A). Subsequent work revealed that
177 when these three pathways are eliminated from the ancestral type evolution proceeds
178 along multiple new pathways (Lind, et al. 2015) resulting in WS types whose
179 fitnesses are indistinguishable from those arising via mutations in Wsp, Aws, or Mws.
180 Parallel evolution of WS involving preferential usage of the Wsp, Aws and Mws
181 pathways is therefore not explained by selection: repeated use of Wsp, Aws and Mws
182 stems from the fact that these pathways are subject to negative regulation and thus,
183 relative to pathways subject to positive regulation, or requiring promoter-activating
184 mutations, gene fusion events, or other rare mutations, present a large mutational
185 target (Lind, et al. 2015).
186

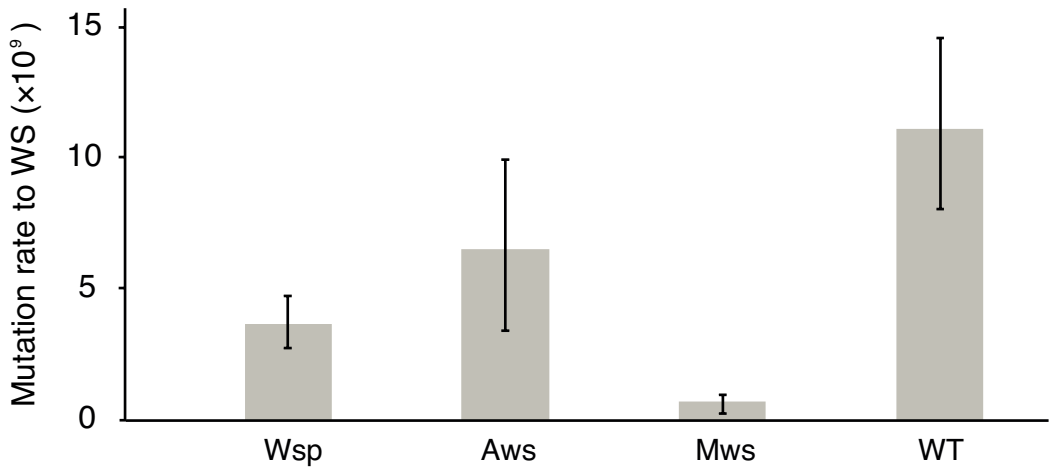
187 Given repeatability of WS evolution, knowledge of the Wsp / Aws / Mws pathways,
188 plus genetic tools for mechanistic investigation — including capacity to obtain WS
189 mutants in the absence of selection — the WS system offers a rare opportunity to
190 explore the extent to which knowledge of the genotype-to-phenotype map can lead
191 mechanistic models for evolutionary forecasting. Our findings show that short-term
192 mechanistic-level predictions of mutational pathways are possible, but also draw
193 attention to challenges that stem from current inability to *a priori* predict locus-
194 specific mutational biases and environment-specific fitness effects.
195

196 ## Results

197 **Obtaining an unbiased measure of pathway-specific mutation rates to WS**
198 Knowledge of the rate at which mutation generates WS types via each of the Wsp,
199 Aws and Mws pathways — unbiased by the effects of selection — provides a
200 benchmark against which the predictive power of null models can be appraised. To
201 achieve such measures we firstly constructed a set of genotypes containing just one of
202 the three focal pathways: PBR721 carries the Wsp pathway but is devoid of Aws and
203 Mws, PBR713 carries the Aws pathway but is devoid of Wsp and Mws, while
204 PBR712 harbours the Mws pathway but is devoid of Wsp and Aws. Into each of
205 these genotypes a promoterless kanamycin resistance gene was incorporated
206 immediately downstream of the promoter of the cellulose-encoding *wss* operon and
207 transcriptionally fused to an otherwise unaffected *wss* operon (Figure 1B).
208

209 In the ancestral SM genotype the cellulose promoter is inactive in shaken King's
210 Medium B (KB) broth (Spiers, et al. 2002) and thus the strain is sensitive to
211 kanamycin. When a WS-causing mutation occurs, the *wss* promoter becomes active

212 resulting in a kanamycin-resistant WS type (Fukami, et al. 2007; McDonald, et al.
213 2011). Individual growth of this set of three genotypes in shaken KB, combined with
214 plating to detect kanamycin-resistant mutants, makes possible a fluctuation assay
215 (Luria and Delbruck 1943; Hall, et al. 2009) from which a direct measure of the rate
216 at which WS mutants arise can be obtained. Importantly, because WS types are
217 maladapted in shaken broth culture, the screen for kanamycin-resistant clones allows
218 the pathway-specific mutation rate to WS to be obtained without the biasing effects of
219 selection for growth at the air-liquid interface (Figure 1B). The results are shown in
220 Figure 2.



221

222 **Figure 2. Mutation rates to WS.** Fluctuation tests were used to estimate the mutation
223 rate to WS for the three common mutational pathways to WS. Error bars represent
224 mean and 95% confidence intervals. All mutation rates are significantly different
225 from all others (t-test $P < 0.05$, see method section for details). Number of replicates n
226 = 200 for Aws and Wsp, n = 400 for Mws, n = 100 for WT.

227

228 The mutation rate was highest for the Aws pathway (6.5×10^{-9}); approximately
229 double that of Wsp (3.7×10^{-9}) and an order of magnitude higher than that of the Mws
230 pathway (0.74×10^{-9}) (Figure 2). The rate at which WS mutants arose from the

231 ancestral genotype in which the three pathways are intact (11.2×10^{-9}) was
232 approximately the sum of the rates for the three pathways (11.0×10^{-9}) confirming
233 that the Wsp, Aws and Mws pathways are the primary routes by which WS types
234 evolve (Lind, et al. 2015). That the Aws pathway has the greatest capacity to
235 generate WS is surprising given the smaller target size (three genes and 2.3 kb
236 compared to seven genes (8.4 kb) in the Wsp pathway).
237
238 **The genotype-to-phenotype map underpinning WS evolution**
239 Much is known about the function and interactions among components of each of the
240 three focal pathways. This knowledge allows development of models that capture the
241 dynamic nature of each pathway and thus allow predictions as to the likelihood that
242 evolution will precede via each of the three mutational routes. An unresolved issue is
243 the extent to which these models match experimental findings. Following a brief
244 description of each pathway we describe the models.
245
246 The 8.4 kb Wsp pathway is a chemotaxis-like system (Goymer, et al. 2006; Guvener
247 and Harwood 2007; Römling, et al. 2013; Micali and Endres 2016) composed of
248 seven genes with the first six genes (*wspA-wspF*) being transcribed as a single unit
249 and the last (*wspR* from its own promoter (Bantinaki, et al. 2007). WspA
250 (PFLU1219) is a methyl-accepting chemotaxis (MCP) protein that forms a complex
251 with the CheW-like scaffolding proteins WspB (PFLU1220) and WspD (PFLU1222).
252 WspA senses environmental stimuli and transmits the information via conformational
253 changes in the WspA/WspB/WspD complex to effect activity of WspE (PFLU1223),
254 a CheA/Y hybrid histidine kinase response regulator. WspE activates both the WspR
255 (PFLU1225) diguanylate cyclase (DGC) and the CheB-like methylesterase WspF

256 (PFLU1224) following transference of an active phosphoryl group. The activity of
257 WspA is modulated by methylation by the constitutively active CheR-like
258 methyltransferase WspC (PFLU1221) that transfers methyl groups to conserved
259 glutamine residues on WspA. The demethylase WspF serves to remove these groups
260 when in the phosphorylated active form. WS mutants are known to arise by
261 mutations in the WspF negative regulator and also in the WspE kinase (McDonald, et
262 al. 2009). In vitro manipulations of WspR that abolish repression of the DGC domain
263 by the response regulator domain are known, but have never been observed to occur
264 in experimental populations (Goymer, et al. 2006).
265
266 The 2.3 kb *aws* operon contains three genes transcribed from a single promoter
267 (*awsXRO*). Homologous genes in *Pseudomonas aeruginosa* (*yfiRNB,* PA1121-1119)
268 have been characterised in detail (Malone, et al. 2010; Malone, et al. 2012; Xu, et al.
269 2016). The outer membrane lipoprotein AwsO (PFLU5209) has an OmpA domain, a
270 signal peptide and binds to peptidoglycan. AwsO is thought to be the sensor whose
271 activity is modulated in response to envelope stress (Malone, et al. 2012). AwsO
272 sequesters the periplasmic protein AwsX (PFLU5211) at the outer membrane. AwsX
273 functions as a negative regulator of the DGC AwsR (PFLU5210) in the inner
274 membrane. Both increased binding of AwsX to AwsO or loss of negative regulation
275 by inactivation of the interaction between AwsX and AwsR can lead to WS
276 (McDonald, et al. 2009; Malone, et al. 2010; Malone, et al. 2012).
277
278 The 3.9 kb *mwsR* gene (PFLU5329) is known as *morA* (PA4601) in *Pseudomonas*
279 *aeruginosa*, and encodes a predicted membrane protein with both a DGC domain that
280 produces c-di-GMP and a phosphodiesterase (PDE) domain that degrades c-di-GMP.

281 Little is known of the molecular details determining its function, but both catalytic
282 domains appear to be active (Phippen, et al. 2014). Deletion of the PDE domain
283 results in a WS phenotype with activity being dependent on a functional DGC domain
284 (McDonald, et al. 2009).
285
286 **Development of mechanistic models predicting mutational routes to WS**
287 If the specific effects of changing each nucleotide (and sets of nucleotides) were
288 known then models for each pathway would not be required. Here, we show how
289 knowledge of genetic architecture can be used to build models that predict the
290 likelihood that mutations generating WS types will arise in a given pathway – and
291 even in specific genes. In the following section we present four null models that
292 incorporate increasing levels of information concerning the genotype-to-phenotype
293 map. The goal of these models is two-fold: firstly to demonstrate that incorporation
294 of knowledge of genetic architecture allows development of models with explanatory
295 value and secondly, to define minimal necessary information for reliable forecasting.
296 The results are summarised in Figure 3, which displays the experimental data from
297 Figure 2 (Figure 3A), along with predictions from each of the null models.

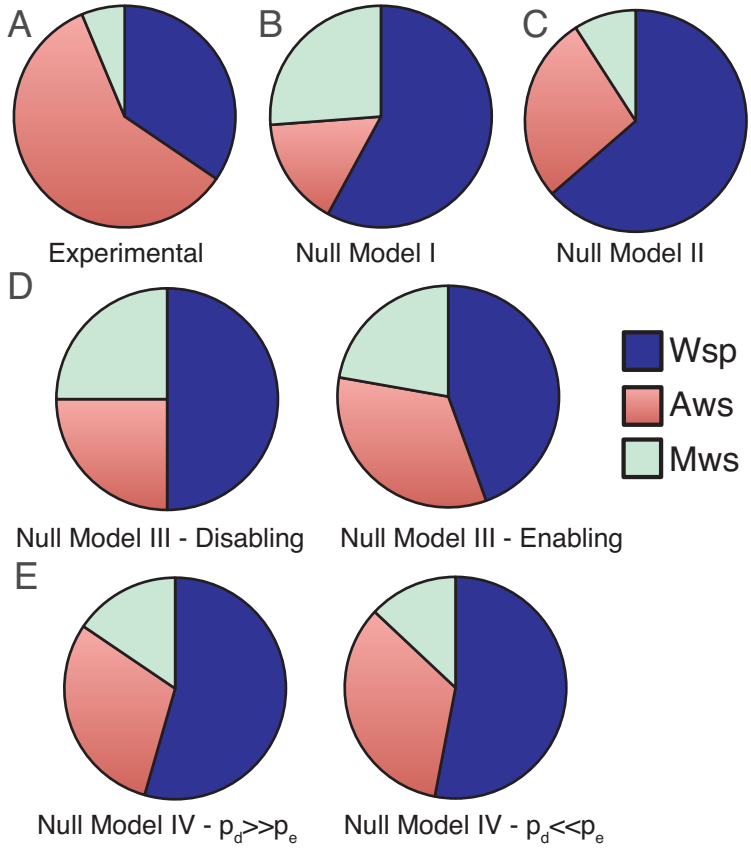


Figure 3. Comparisons of experimental data and null model predictions for the use of the Wsp, Aws and Mws pathways. (A) Experimental data from Figure 2. **(B)** Null Model I: number of base pairs as proxy for mutational target size. **(C)** Null Model II: number of genes as a proxy for mutational target size. **(D)** Null Model III: function and interactions between components determine mutational target size. To indicate the range of possible states we consider the situation with enabling mutations only (left hand panel) and disabling mutations only (right hand panel). **(E)** Null Model IV: as per null Model III, but interactions have both pleiotropic and continuous effects. To indicate the range of possible states we consider the probability of disabling mutational effects (p_d) to be 10 times more common than enabling changes (p_e), $p_d = 0.001$, $p_e = 0.0001$ (left hand panel) the probability of enabling mutational effects (p_e) to be 10 times more common than disabling changes (p_d), $p_d = 0.0001$, $p_e = 0.001$.

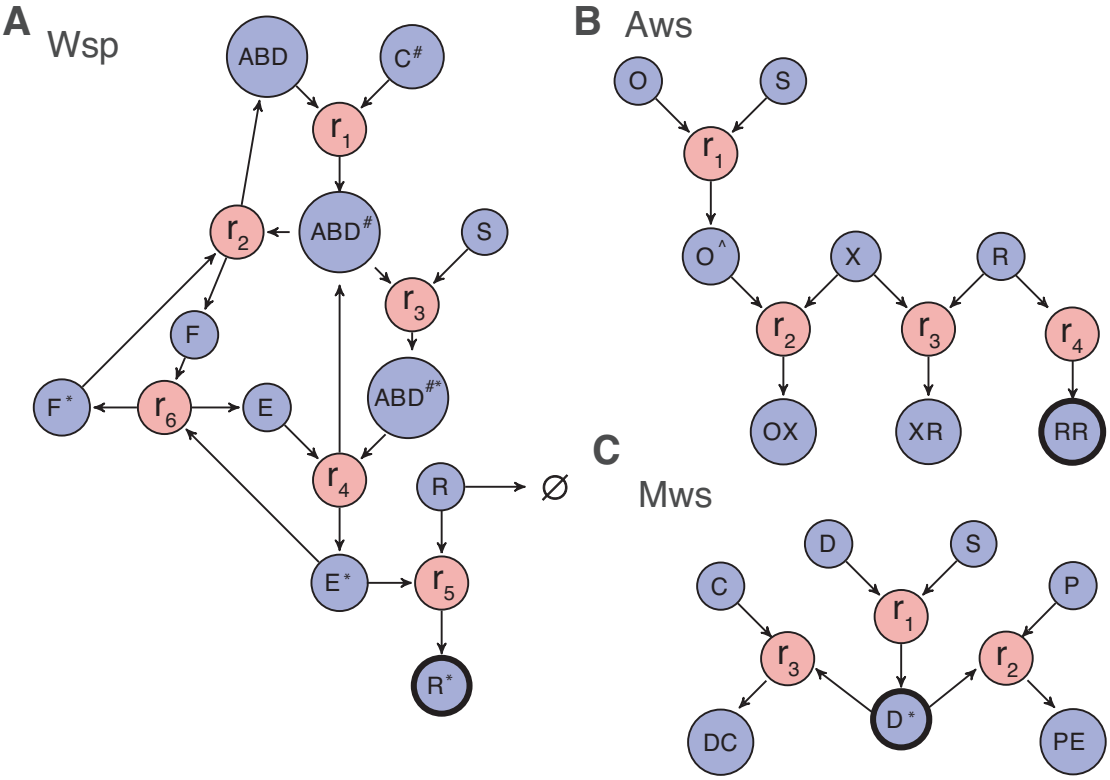
312

313 Null Model I is intentionally naïve. It uses sequence length as a proxy for mutational
314 target size, but ignores genetic organisation, function of predicted proteins and
315 interaction among proteins. The model assumes that mutational target size is
316 proportional to the number of nucleotides at a given locus and thus the probability
317 that a given pathway is used to generate WS relative to another is simply the ratio of
318 the probability of generating WS for the two focal pathways. For any one pathway
319 the probability that a mutation generates WS is given by $1-(1-p)^n$, where p is the
320 probability of a mutation at a nucleotide and n is the number of nucleotides in the
321 pathway. If the mutation probability is low such that the expected number of
322 mutations in a pathway is below 1, that is, $np<<1$ then the binomial approximation
323 can be used: $1-(1-p)^n = np$. Thus the probability that evolution follows the Wsp
324 pathway over the Aws pathway is: $8400\ p_1 / 2300\ p_2 = 3.65\ p_1/p_2$, where p_1 and p_2 are
325 the mutation rates for each pathway. Assuming equivalency of mutation rate, $p_1 = p_2$,
326 evolution is predicted to proceed via the Wsp pathway 3.65 times more often than via
327 Aws, with evolution predicted to proceed via Mws 1.65 times more often than Aws.
328 Comparison with experimental data shows a departure both in terms of the priority of
329 pathways used by evolution and the frequency of pathway usage (Figure 3A versus
330 3B).

331

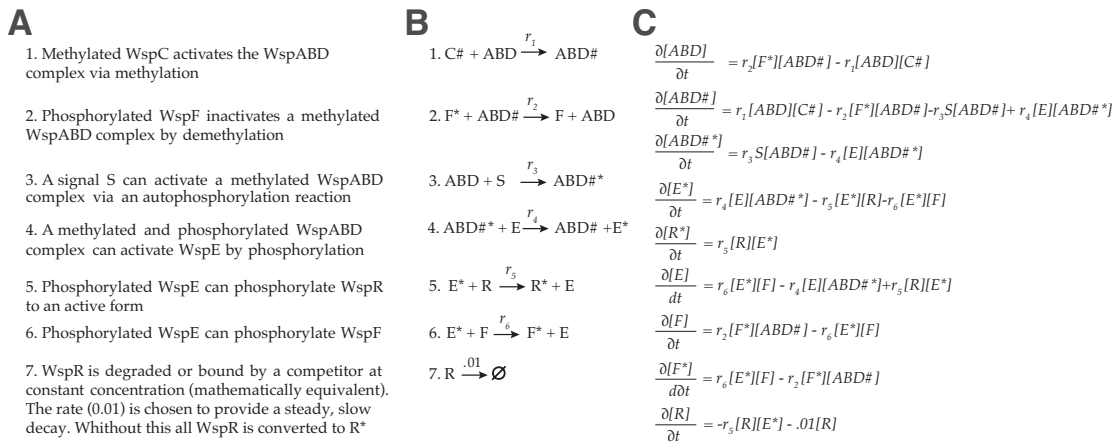
332 Null Model II builds on Model I but only in a marginal sense. It recognises that
333 nucleotides defining loci of interest are organised into genes, and therefore adopts
334 gene number as a proxy for mutational target size. As above, the probability that a
335 pathway is used is $1-(1-p)^n = np$, but in this instance n is the number of genes. Model
336 II predicts that mutations in Wsp generate WS types 2.33 times more often than

337 mutations in Aws, with mutations in Mws being 3-fold less likely to generate WS
338 compared to mutations in Aws (Figure 3C). This marginal adjustment makes little
339 difference to the fit between experimental data and predictions.
340
341 Past work has shown the explanatory value of information that comes from
342 knowledge of gene (protein) function and interactions (McDonald, et al. 2009; Lind,
343 et al. 2015). These form the basis of Null Model III. The relevant functions and
344 interactions are depicted in Figure 4 as reaction diagrams that reflect how changes in
345 different interactions affect the production of WS types. Organizing interactions
346 within each pathway according to reactions has the advantage that it allows for a
347 standard mathematical description of the biochemical dynamics using differential
348 equations (see Figure 4 and Figure 4 - figure supplements 1 to 3,). An additional
349 advantage is that such an approach allows ready incorporation of new experimental
350 data including interaction type and interaction strength.



351

352 **Figure 4. Reaction diagrams for null models III and IV.** Reaction diagrams show
353 the interactions governing production of a wrinkly spreader in (**A**) Wsp, (**B**) Aws and
354 (**C**) Mws pathways. The blue circles indicate molecular components, the red circles
355 indicate reactions, and arrows indicate which molecular components are reactants and
356 which are products. Full descriptions of the models including all reactions and the
357 resulting systems of differential equations are found in Figure 4 figure supplement 1
358 for Wsp, Figure 4 figure supplement 2 for Aws and Figure 4 figure supplement 3 for
359 Mws. The black edged circles are the reporter proteins, which is the activated forms
360 of the diguanylate cyclases that directly determine whether a pathway generates a
361 wrinkly spreader. Increased production of any reporter leads to increased c-di-GMP
362 concentration causing increased cellulose production and thereby a wrinkly spreader.
363



364

365 **Figure 4 - figure supplement 1. Wsp model (A)** Description of functional
366 interactions **(B)** Description of molecular reaction **(C)** Differential equations
367 describing the dynamics of the Wsp pathway. The activity of WspA is modulated by
368 methylation, where it is activated by the CheR-like methyltransferase WspC
369 (PFLU1221). The CheB-like methylesterase WspF (PFLU1224) functions as a
370 negative regulator. Modulation of WspR activity through changes in oligomerization
371 state and clustering is not explicitly included in the model (De, et al. 2008;

Huangyutitham, et al. 2013), but can be interpreted as changes in the rate of WspR activation.

A

1. A signal S binds to AwsO and activates it
2. Activated AwsO binds to AwsX to form a complex that sequesters AwsX
3. AwsX binds AwsR to form a complex that sequesters AwsX
4. AwsR dimerizes (active form)

B

1. $S + O \xrightarrow{r_1} O^\wedge$
2. $X + O^\wedge \xrightarrow{r_2} OX$
3. $X + R \xrightarrow{r_3} XR$
4. $R + R \xrightarrow{r_4} RR$

C

$$\frac{\partial[X]}{\partial t} = -r_2[X][O^\wedge] - r_3[X][R]$$

$$\frac{\partial[XR]}{\partial t} = r_3[X][R]$$

$$\frac{\partial[O]}{\partial t} = -r_1 S[O]$$

$$\frac{\partial[O^\wedge]}{\partial t} = r_1 S[O] - r_2[X][O^\wedge]$$

$$\frac{\partial[OX]}{\partial t} = r_2[X][O^\wedge]$$

$$\frac{\partial[R]}{\partial t} = -r_3[X][R] - r_4[R][R]$$

$$\frac{\partial[RR]}{\partial t} = r_4[R][R]$$

Figure 4 - figure supplement 2. Aws model (A) Description of functional interactions **(B)** Description of molecular reaction **(C)** Differential equations describing the dynamics of the Aws pathway. Release of AwsX mediated repression results in a conformational shift that in the model is represented as formation of an active dimer.

A

1. A signal S binds to DGC domain and activates it
2. Activated DGC interacts with PDE to inactivate it
3. Activated DGC interacts with C and to form alternative complex DC

B

1. $S + D \xrightarrow{r_1} D^*$
2. $D^* + P \xrightarrow{r_2} DP$
3. $D^* + C \xrightarrow{r_3} DC$

C

$$\frac{\partial[D]}{\partial t} = -r_1[D]S$$

$$\frac{\partial[D^*]}{\partial t} = r_1[D]S - r_2[D^*][P] - r_3[D^*][C]$$

$$\frac{\partial[P]}{\partial t} = -r_2[D^*][P]$$

$$\frac{\partial[DP]}{\partial t} = r_2[D^*][P]$$

$$\frac{\partial[C]}{\partial t} = -r_3[D^*][C]$$

$$\frac{\partial[DC]}{\partial t} = r_3[D^*][C]$$

384 **Figure 4 - figure supplement 3. Mws model (A)** Description of functional
385 interactions **(B)** Description of molecular reactions **(C)** Differential equations
386 describing the dynamics of the Aws pathway
387
388 One immediate consequence of this approach is that production of WS types is
389 described entirely through biochemical reactions. Knowing whether a WS-type is
390 generated by a particular mutation amounts to determining which reaction rates are
391 altered by the mutation and whether those changes affect the likelihood of producing
392 WS types. This leads to an approach where mutations in components can be classified
393 according to their effects on reaction rates: *enabling* mutations increase reaction rates,
394 whereas *disabling* mutations decrease reaction rates.
395
396 Based on the mathematical models for the pathways, each reaction rate can be altered
397 (either increased or decreased) to generate higher levels of the key WS effector (the
398 active DGC). For example, in the Wsp pathway, if reactions 2 or 6 experience a
399 disabling mutation, or alternatively, if any of the reactions 1, 3, 4, or 5 experience an
400 enabling mutation, then the level of activated (phosphorylated) WspR is increased.
401 For now, it is assumed that increased activity of the relevant DGC generates a WS
402 type. Reactions whose enabling / disabling mutations increase the amount of WS
403 effector are termed enabling / disabling reactions.
404
405 By comparing the number of disabling and enabling reactions in different pathways it
406 is possible to calculate the relative likelihood that evolution uses each mutational
407 pathway. The Wsp pathway encompasses two disabling and four enabling reactions,
408 whereas the Aws pathway is defined by one disabling and three enabling reactions.

409 Mws encompasses a single disabling reaction and two enabling reactions. If disabling
410 mutations are as likely as enabling mutations, then the likelihood that a pathway will
411 be used by evolution is simply the ratio of the total number of enabling and disabling
412 reactions: 6:4:3 (Wsp:Aws:Mws). If instead, disabling mutations are much more
413 likely than enabling mutations, then enabling reactions can be ignored and the
414 likelihood that a pathway will be used by evolution becomes simply the ratio of the
415 number of disabling reactions: 2:1:1 (Wsp:Aws:Mws) (Figure 3D left hand panel). If
416 the reverse is true, then the likelihood of use is a ratio of the number of enabling
417 reactions: 4:3:2 (Wsp:Aws:Mws) (Figure 3D right hand panel). In all cases, mutations
418 in the Wsp pathway are predicted to be 1.3-2 times more likely to generate WS than
419 mutations in the Aws pathway, with the Aws pathway being the target of evolution 1-
420 1.5 times more often than the Mws pathway. Although still inflating the importance
421 of the Wsp pathway relative to Aws, Mws is more prominent than in null Model II.
422
423 Null Model IV takes a further step toward mechanistic accuracy by endowing
424 interactions with pleiotropic and continuous effects. This removes two simplifying
425 assumptions of null Model III likely to limit predictive power. In Model III mutations
426 affect one reaction at a time, but mutations in certain components, for example *wspA*,
427 affect more than a single reaction (WspA appears in reactions 1-4). Additionally, in
428 Model III, changes in reaction rate are assumed to be binary, however reactions may
429 have a range of effects on the evolution of WS types.
430
431 To accommodate pleiotropic effects, null Model IV systematically considers all
432 combinations of enabling and disabling changes to reaction rates and determines the
433 likelihood that a WS type is generated. An example of one set of the possible

434 mutations (m_i) in Wsp is 1, −1, 0, 0, 0, 0 (an increase in r_1, a decrease in r_2, but no
435 change in r_3, r_4, r_5, or r_6 (Figure 3A)). Since the Wsp pathway has six reaction rates
436 this amounts to 3^6 or 729 total combinations. However, note that reaction 3 does not
437 share any reactants or products with reactions 5 or 6. Thus, mutations such as 0, 0, 1,
438 0,1, 0 or 0, 0,1, 0, 0, 1 are not considered because they require mutations in two
439 separate genetic components.
440
441 To accommodate a range of effects null Model IV simulates enabling / disabling
442 changes of different magnitudes and determines the resulting effect on the respective
443 effector DGC (see Materials and Methods). Briefly, the approach addresses the lack
444 of information concerning biochemical reaction rates and molecular concentrations in
445 the mathematical models describing WS-producing pathway dynamics. By repeatedly
446 sampling from the space of all possible reaction rates, initial concentrations, and
447 magnitudes of effects, this approach computes the probability that a particular set of
448 mutations (m_i), e.g. 1, −1, 0, 0, 0, 0, results in a wrinkly spreader. This probability is
449 represented as the conditional probability $P\ (WS\ |m_i \in Wsp)$, which motivates a
450 Bayesian formulation to compare the relative probability that the different pathways
451 produce WS. To this end, the probability that a particular pathway will be used is
452 decomposed into two terms: the probability that a particular set of mutations (m_i)
453 occurs in Wsp (or Aws, or Mws) represented as $P\ (m_i \in Wsp)$ and the probability that
454 those mutations give rise to a wrinkly spreader represented as $P\ (WS\ |m_i \in Wsp)$ (or
455 Aws, or Mws).
456
457 $$P(WS \cap m \cap Wsp) = \sum_i P(WS|m_i \in Wsp)P(m_i \in Wsp) \qquad (1)$$
458

459 To estimate $P\ (m_i \in Wsp)$ we assume fixed probabilities of enabling and disabling
460 changes and compute the product. Thus, the probability of $m_i = 1, -1, 0, 0, 0, 0$ is
461 $p_e p_d (1 - p_e - p_d)^4$, where p_e is the probability of a mutation with an enabling effect
462 and p_d is the probability of a mutation with a disabling effect. Recognising the value
463 of accommodating the possibility of localised mutational bias we note that p_e and p_d
464 can be adjusted for the affected reactants. The second term, $P\ (WS\ |m_i \in Wsp)$, relies
465 on our sampling methodology and describes the probability that a set of
466 disabling/enabling changes of different magnitudes will yield a WS type (see
467 Materials and Methods).
468
469 Despite the mechanistic advances incorporated into null Model IV the Wsp pathway
470 is still predicted to be the pathway most commonly used by evolution. The extreme
471 cases in which disabling mutations are more probable than enabling mutations (and
472 vice versa) are shown in Figure 3E. The results of simulations for the full range of p_d
473 and p_e values are shown in Figure 5. Figure 5A shows that the Wsp pathway is
474 predicted to be the target of mutation 1.5 - 1.9 times more often than the Aws
475 pathway while Figure 5B shows that the Mws pathway is predicted to be the target of
476 mutation 0.4 - 0.5 times less often that the Aws pathway. While these results agree
477 with the experimental data showing Mws to be least likely pathway to be followed,
478 the predictions are at odds with the data showing WS types to be twice as likely to
479 arise from mutation in Aws, versus Wsp. The causes of this discrepancy are
480 accounted for in the next section "analysis of mutants reveals sources of mutational
481 bias".

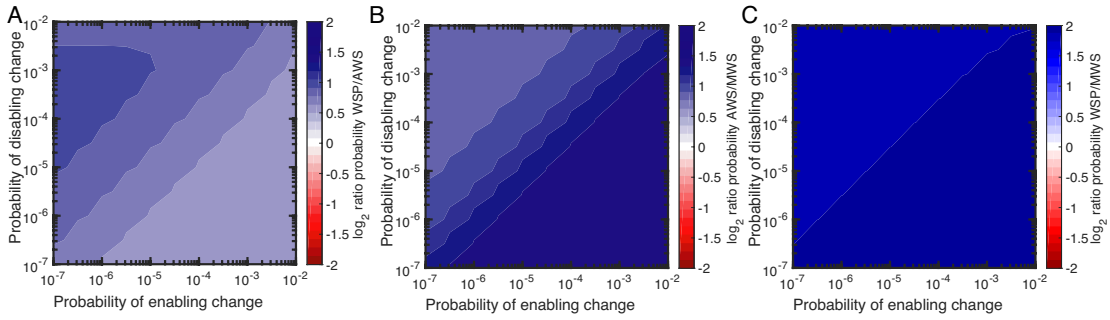


Figure 5. Null Model IV predictions of the probability of using the Wsp, Aws and Mws pathways (A) Probability of Wsp relative to Aws. (**B**) Probability of Mws relative to Aws. **(C)** Probability of Mws relative to Wsp**.** The relative contributions of individual reactions rates are available in Figure 5 - figure supplement 1 for Wsp and Figure 5 - figure supplement 2 for Aws.

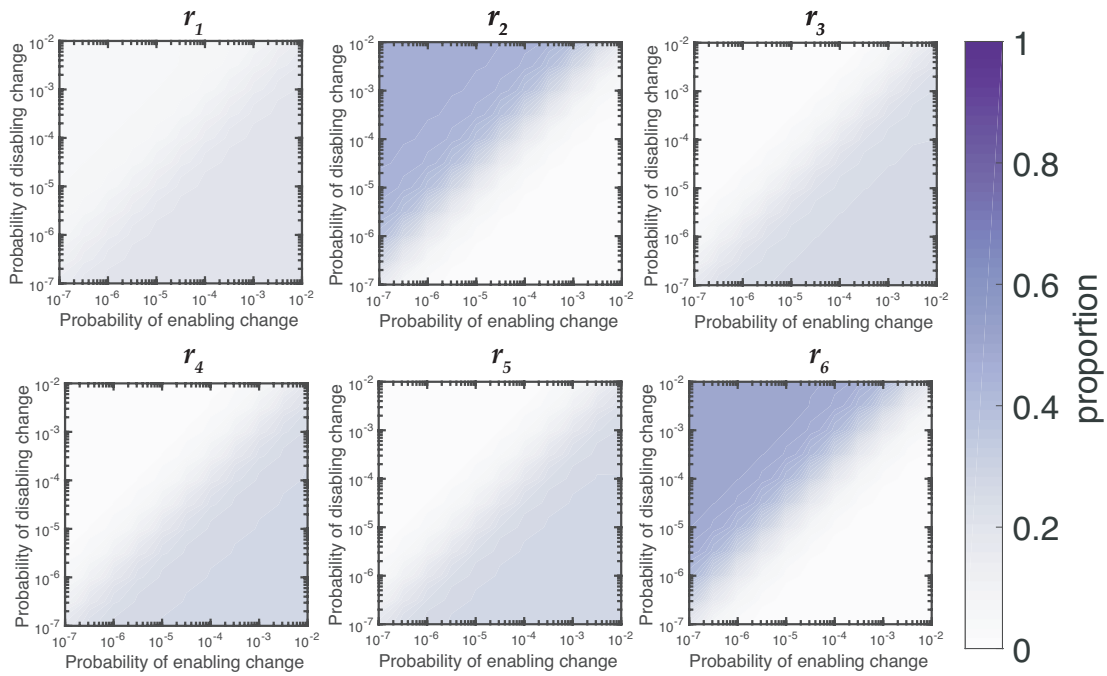
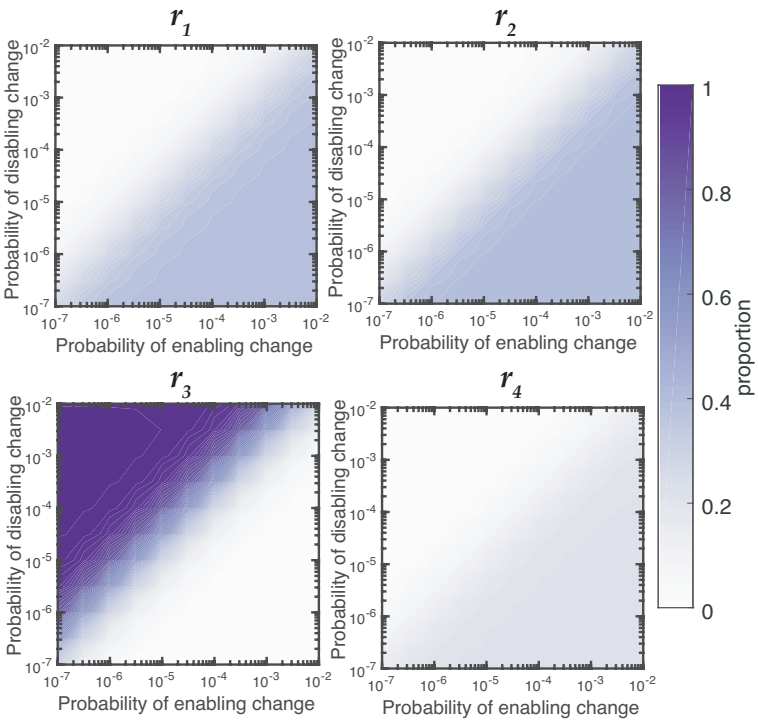


Figure 5 - figure supplement 1. Relative contributions of reaction rates for Wsp for null Model IV. Proportion of WS mutants that are produced by changes to each reaction rates is dependent on the rates of enabling and enabling mutations. When the

493 rate of disabling mutations is much higher than the rate of enabling mutations the

494 majority of WS mutations in Wsp are produced by changes to r_2 and r_6.

495



496

497 **Figure 5 - figure supplement 2. Relative contributions of reaction rates for Aws**

498 **for null Model IV.** Proportion of WS mutants that are produced by changes to each

499 reaction rates is dependent on the rates of enabling and enabling mutations. When the

500 rate of disabling mutations is much higher than the rate of enabling mutations the

501 majority of WS mutations in Wsp are produced by changes to r_3.

502

503 The Bayesian approach of null Model IV has additional utility: it allows prediction of

504 genes likely to be affected by mutation. Predictions as to specific genetic targets

505 come from appraisal of the relative importance of each reaction in terms of generating

506 WS types (Figure 5 – supplement 1 and 2). While it is recognised that a majority of

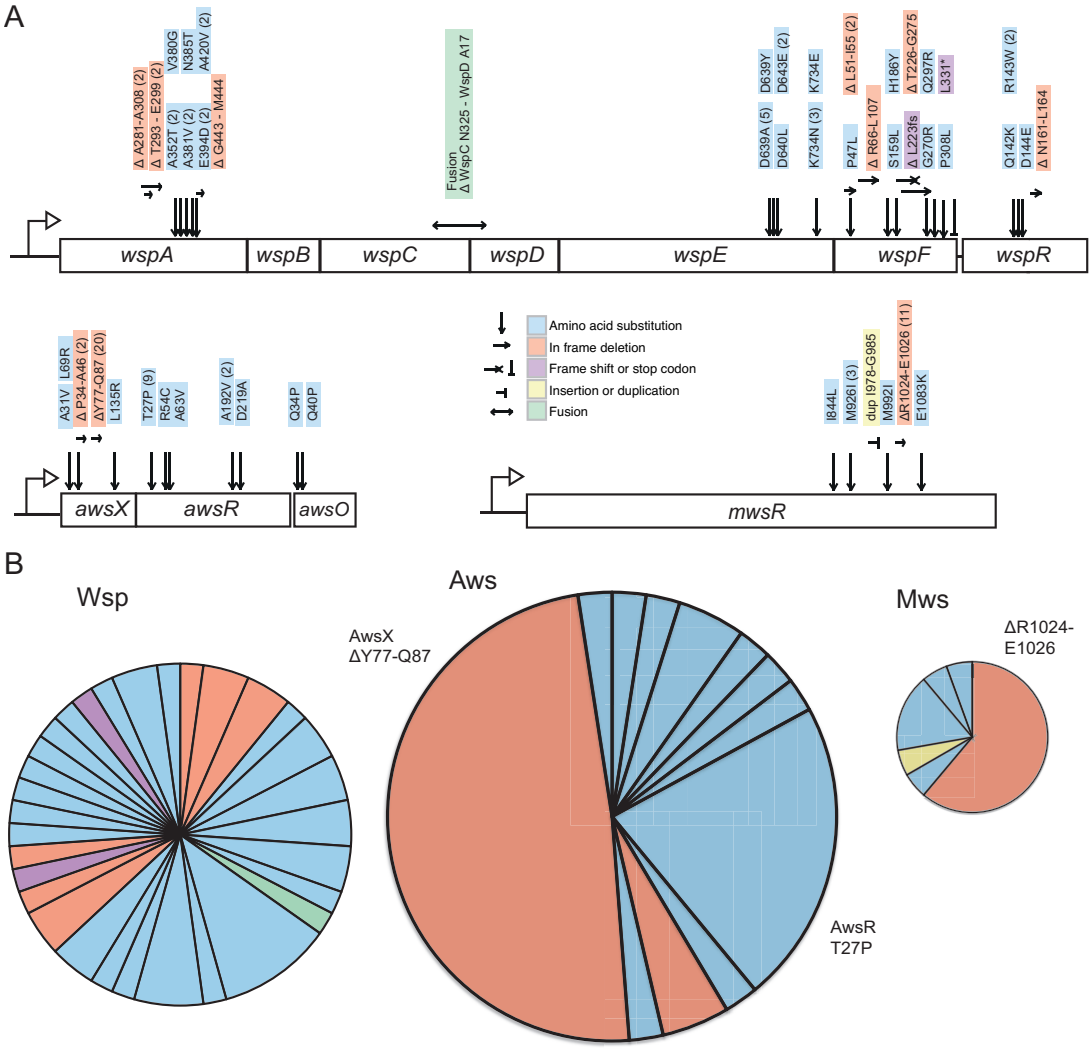
507 WS mutations arise from defects in negative regulators of DGCs, such as WspF and

508 AwsX (McDonald, et al. 2009; Lind, et al. 2015), further predictions are possible

509 based on impacts of alterations in gene function on reaction rates. For example, with
510 reference to the Wsp pathway (Figure 5 – figure supplement 1), there are two
511 reactions (2 and 6) that affect WspF function: r_2 describes the rate of removal of
512 methyl groups from the signalling complex and r_6 the rate at which WspF is activated
513 by transfer of active phosphoryl groups from the WspE kinase. Loss-of-function
514 (disabling) mutations being much more common than gain-of-function (enabling)
515 mutations means that both WspF and WspE are likely targets. The null model
516 predicts that in the area of parameter space in which Wsp is most likely compared to
517 Aws, 45% of the time WS will be generated when the second reaction, r_2, is altered
518 (Figure 5 – figure supplement 1). The same is true for reaction r_6. Thus the presence
519 of a negative regulator is predicted to extend the mutational target size well beyond
520 the gene itself. This is also true for Aws, where r_3 is the main contributor to the WS
521 phenotype in the case where disabling change is more common than enabling change.
522 Here mutations are predicted not only in the negative regulator AwsX, but also in the
523 interacting region of the DGC AwsR (Figure 5 – figure supplement 2).
524
525 Loss-of-function mutations in negative regulators and their interacting partners are
526 not the only predicted targets. For Wsp r_1, r_3, r_4, and r_5 are altered approximately 5%
527 of the time in the parameter region where disabling mutations are more common than
528 enabling mutations, but contribute more when the rate of enabling mutations is
529 increased (Figure 5 – figure supplement 1). Enabling mutations based on null Model
530 IV are likely to be found in WspC (increasing r_1), WspABD (increasing r_3),
531 WspABD/WspE (increasing r_4) and WspR (increasing r_5) (Figure 4A). For Aws,
532 enabling mutations are predicted to increase reaction r_1 by mutations causing
533 constitutive activation of AwsO, r_2 increasing binding of AwsO and AwsX and r_4

534 increasing formation of the active AwsR dimer (Figure 4B, Figure 5 – figure
535 supplement 2).
536
537 In summary, high rates of WS mutations are predicted for *wspF*, *wspE, wspA, awsX*
538 and *awsR* with lower rates for *wspC*, *wspR* and *awsO*. Several of these predictions sit
539 in accord with previous experimental observations, however, notable are predictions
540 that evolution might also target *wspA* and *wspR*, two genes that have not previously
541 been identified as mutational causes of WS types (McDonald, et al. 2009).
542
543 **Analysis of mutants reveals sources of mutational bias**
544 There are several reasons why predictions from the models might be out of kilter with
545 experimental data on mutation rates. We firstly looked to the distribution of WS
546 generating mutations among the 109 mutants collected during the course of the
547 fluctuation assays. Of the 109 mutants, 105 harboured a mutation in *wsp* (46
548 mutants), *aws* (41 mutants) or *mws* (18 mutants) (Figure 6A, Figure 6 – source data
549 1). The remaining four had mutations in previously described rare pathways
550 (PFLU0085, PFLU0183), again confirming that these non-focal pathways produce
551 just a fraction of the total set of WS mutants (Lind, et al. 2015).

552



553

554 **Figure 6. Mutational targets.** (**A**) 105 independent mutations in the *wsp* (n=46), *aws*

555 (n=41) and *mws* (n=18) operons were identified. Numbers of independent mutants are

556 shown in brackets. Full details on the mutations are available in Figure 6 – source

557 data 1. (**B**) Diversity of mutations with area proportional to mutation rate (Figure 2).

558 Two mutations (AwsX ΔY77-Q87 and AwsR T27P) contribute 41% of all mutations

559 to WS suggesting that these are mutational hot spots.

560

561 The distribution of mutations for each of the three pathways is indicative of bias. As

562 shown in Figure 6B, almost 29% of all WS-causing mutations (adjusted for

563 differences in mutation rates between the three pathways) were due to an identical 33

564 base pair in-frame deletion in *awsX* (Δt229-g261, ΔY77-Q87), while a further 13 %
565 were due to an identical mutation (79 a->c, T27P) in *awsR*. At least 41 different
566 mutations in Aws lead to WS: if mutation rates were equal for these sites the
567 probability of observing 20 identical mutations would be extremely small. In fact 10
568 million random samplings from the observed distribution of mutations failed to
569 recover this bias. While the Wsp pathway also contains sites that were mutated more
570 than once (six positions were mutated twice, one site three times and one five times),
571 sources of mutational bias in Wsp were less evident than in Aws (Figure 6B).
572
573 **Incorporating mutational bias into null models**
574 The mathematical models presented above assumed no mutational bias, but the null
575 models can be modified to incorporate such bias. With focus on the AwsX hotspot,
576 we show this for models II-IV.
577
578 Null Model II: The probability of a mutation at the hotspot is kp where $k>1$ and p is
579 the probability of a mutation in non-hotspot locations. The probability that a pathway
580 is used is: $kp + (1-kp)(1-(1-p)^{(n-1)})$ where kp is the probability of a mutation at the
581 hotspot and $(1-kp)(1-(1-p^{)(n-1)})$ is the probability there is no mutation at the hotspot,
582 but that there is a mutation elsewhere in the pathway. Using the binomial
583 approximation, the probability a given pathway is used is: $kp + (1-kp)(n-1)p$, which
584 for Aws is $kp+(1-kp)2p$. Figure 7A shows that if the mutation rate at the hotspot
585 increases 10-fold ($k = 10.3$) then Aws is used 1.8 times more often than Wsp. While
586 this matches experimental data, the same 10-fold increase applied to Mws results in
587 mismatch. Only with a reduced mutation rate at the same locus ($k – 6.8$) do predicted
588 and observed data agree (Figure 7B).

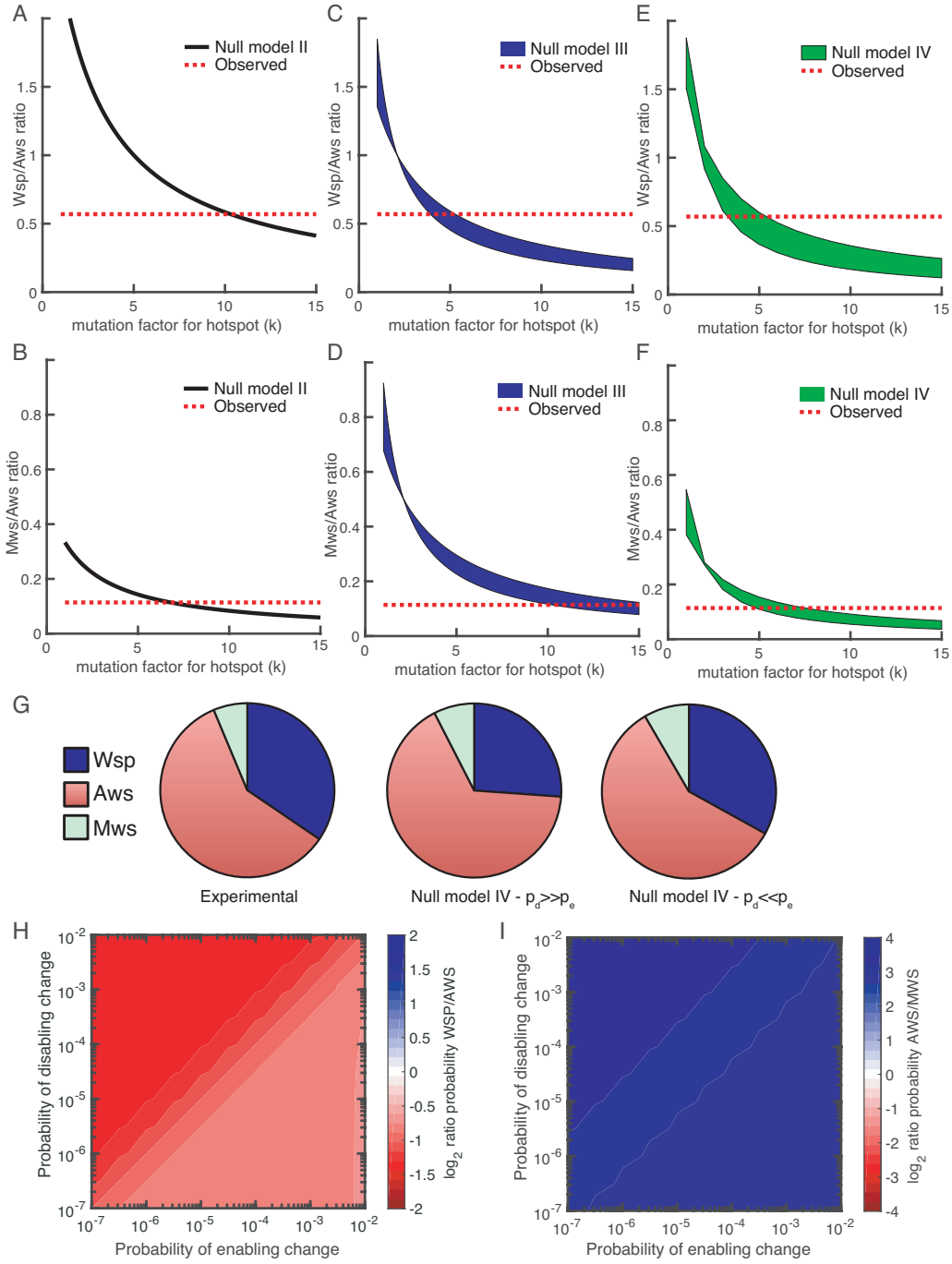


Figure 7. Revised models including mutational hot spot. Null Models II (**A**, **B**), III (**C**, **D**) and IV (**E**, **F**) can be revised to take into account the mutational hot spot in AwsX by including a mutation factor (*k*). Each plot (**A-F**) shows the predicted ratios of Wsp/Aws or Aws/Mws compared to the experimental results (dashed red line). For null models III and IV a range is shown that depends on the probabilities of enabling (p_e) and disabling (p_d) changes. (**G**) Comparison of experimental data (left) and

revised predictions from null model IV (middle, right), incorporating a five-fold increase in p_e and p_d for r_2 and r_3 in the Aws differential equation system. Middle pie chart use p_d=0.001, p_e=0.0001 and right pie chart p_d=0.0001, p_e=0.001 to allow comparison to Figure 3F and 3G. (**H**) Null Model IV predicted probability of Wsp relative to Aws with a hotspot in AwsX that increases the mutation rate five-fold for p_e and p_d for r_2 and r_3 in the Aws system of differential equation (**I**) Null Model IV predicted probability of Aws relative to Mws with a hotspot in AwsX that increase the mutation rate five-fold for p_e and p_d for r_2 and r_3 in the Aws system of differential equation. Figure 7 - figure supplement 1 shows the relative contribution of each reaction rate in the Aws network to the production of WS for the revised null model IV with a 5 times increase for p_e and p_d for r_2 and r_3.

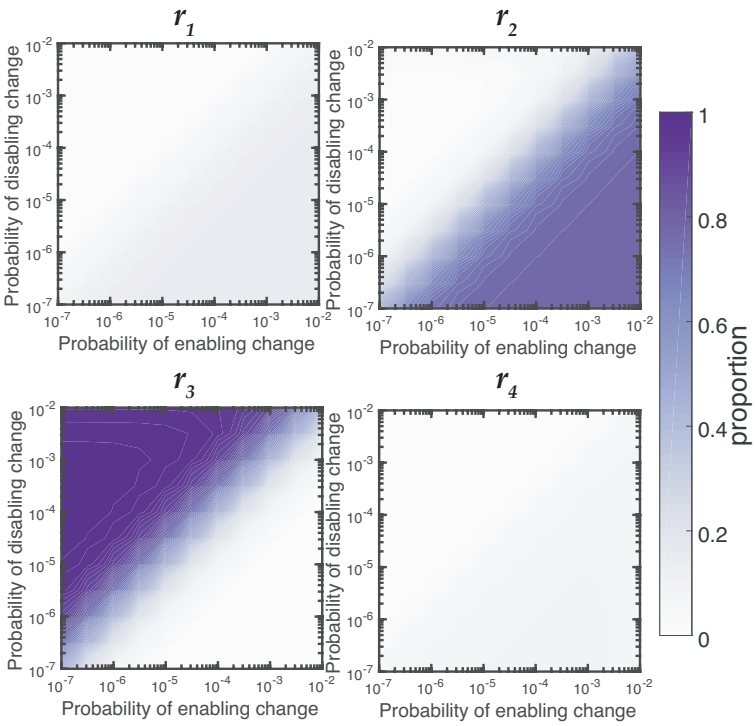


Figure 7 - figure supplement 1. Relative contributions of reaction rates for Aws for null Model IV with a five-fold increase in mutation factor for r_2 and r_3. Proportion of WS mutants that are produced by changes to each reaction rates is

612 dependent on the rates of enabling and enabling mutations. When the rate of disabling
613 mutations is much higher than the rate of enabling mutations the majority of WS
614 mutations in Wsp are produced by changes to r_3.
615
616 Null Model III: The mutational hotspot in the Aws pathway occurs in AwsX, which is
617 involved in both an enabling and a disabling reaction. If the hotspot increases the rate
618 of a disabling reaction by a factor k where $k>1$, then the probability the pathway is
619 used is:
620
621 $k\,p_d + (1-k\,p_d)\,(1-(1-p_d)^{(n_d-1)}) + (1-k\,p_d)\,(1-p_d)^{(n_d-1)}\,(1-(1-p_e)^{n_e})$,
622
623 which is the sum of probabilities of the events the hotspot is used, the hotspot is not
624 used but a disabling change occurs, and the hotspot is not used but only an enabling
625 change occurs. Figure 7C shows that if k is between 4 and 5.3, then there is a good fit
626 to experimental data. The actual value depends on the probabilities of
627 enabling/disabling change and which reaction has the hotspot. For Mws the
628 mutational hotspot must have a factor between 10 and 15 (Figure 7D).
629
630 Null Model IV: The mutational hotspot in *awsX* affects reactions rates r_2 and r_3 in the
631 Aws system of differential equations (Figure 4B, Figure 4 – figure supplement 2). If
632 the mutation factor is between 3.3-5.6 for the Wsp / Aws comparison (Figure 7E) or
633 4.8-7.6 for the Mws / Aws comparison (Figure 7F), then predictions match observed
634 data. Thus Model IV is the only one where the same five-fold change in the
635 probability of enabling / disabling change predicts observed data for all three
636 mutational pathways (Figure 7G). This model, updated with hotspot data, predicts

637 that the Aws pathway is more likely to generate WS types than Wsp for all probability
638 values (see Figure 7H). If disabling changes are more probable than enabling changes,
639 then r_3 dominates the predicted rate, whereas when enabling changes are more
640 common r_2 is the major contributor (Figure 7 – figure supplement 1). Knowledge of
641 the functional interaction between AwsX and AwsR allows interpretation of the likely
642 effect of the hotspot mutations in AwsX and AwsR as disabling change in r_3 (Table
643 1).
644
645 **Analysis of mutants reveals mutational targets and effects**
646 Null Model IV of the Wsp, Aws, and Mws pathways allow estimation of the relative
647 probability that a WS is produced by enabling or disabling changes in a certain
648 reaction rate. This means that in order to evaluate the success of our predictions, the
649 105 WS mutations must be connected to their likely effects on reaction rates. For
650 some mutations this is straightforward. If a mutation completely disables the negative
651 regulator WspF, this will reduce r_2 and r_6, thereby producing a wrinkly spreader. In
652 other cases interpretations of likely effects are more difficult and requires knowledge
653 of the molecular functions of the proteins and their interactions, which needs to be
654 obtained from literature, bioinformatics and structural modelling. For example,
655 mutations in WspA can potentially produce WS by disruption of the interaction with
656 WspF (reducing r_2), increase methylation by WspC (increasing r_1), increase signalling
657 rate (r_3) or increase phosphorylation of WspE (r_4). A summary of the interpretation of
658 the molecular effects of mutations and how they are connected to reaction rates is
659 available in Table 1 with a more detailed description in the supplementary text.
660

Table 1. Connection of mutational effects to reaction rates

Protein	Mutation class	Proposed molecular effect	Proposed reaction rate effect	References
WspA	Amino acid substitutions 352-420	trimer-of-dimer formation, localisation of Wsp clusters, interaction with WspD	Increase r_3	(O'Connor, et al. 2012{Griswold, 2002 #71)
WspA	Deletions A281-A308	Disrupt demethylation by WspF	Decrease r_2	(Rice and Dahlquist 1991)
WspC/D	Fusion of WspC and WspD	Increased methylation by WspC, blocking access of WspF	Increase r_1, decrease r_2	
WspE	Response regulatory domain phosphorylation site	Disrupt phosphorylation of WspF	Decrease r_6	Homology model
WspF	Any disabling mutation	Loss of negative regulation by WspF by demethylation of WspA	Decrease r_2	(Bantinaki, et al. 2007; McDonald, et al. 2009)
WspR	Amino acid substitutions and small deletion in linker region between response regulator and DGC domain	Constitutive activation of DGC without phosphorylation. Effects on subcellular clustering or multimeric state.	Increase r_5	(Goymer, et al. 2006; De, et al. 2009; Huangyutitham, et al. 2013)
AwsX	In frame loss of function mutations	Loss of negative regulation of AwsR	Decrease r_3	(Malone, et al. 2012)
AwsR	Amino acid substitutions transmembrane helix, periplasmic domain	Disruption of interaction with AwsX	Decrease r_3	(Malone, et al. 2012)
AwsR	Amino acid substitutions HAMP linker domain	Changes in dimerization, packing of HAMP domains causing constitutive activation	Increase r_4	(Parkinson 2010; Malone, et al. 2012)

AwsO	Amino acid substitutions in between signal peptide and OmpA domain	Constitutive activation and sequestering of AwsX	Increase r_2	(McDonald, et al. 2009; Malone, et al. 2012; Xu, et al. 2016)

MwsR	Amino acid substitutions and small in frame deletions in interface between EAL and DGC domains	Changes to interdomain interaction or dimerization leading to constitutive activation of DGC function	Decrease r_2, increase r_1	Homology model, (McDonald, et al. 2009)
MwsR	Amino acid substitutions near DGC active site	Loss of feedback regulation or changes to interdomain interaction	Decrease r_2, increase r_1	Homology model, (McDonald, et al. 2009)

661

662

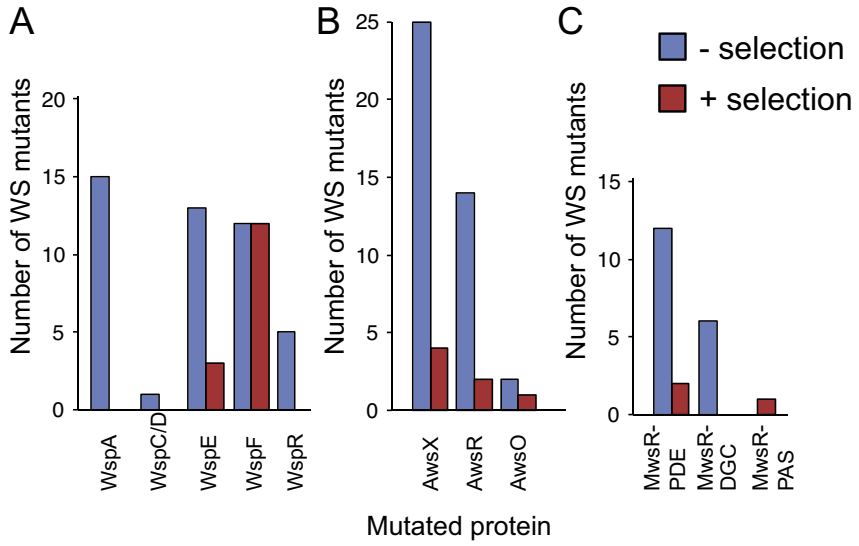
663 Mutations were found in five genes in the Wsp pathway. The majority of the
664 mutations were found in the negative regulator WspF or its interacting partners WspE
665 and WspA. These mutations likely reduced reaction rates of r_2 and r_6 in accord with
666 predictions of null Model IV. The remaining mutations in WspC and WspR were less
667 common, in accord with predictions and likely due to increases in reaction rates r_1 and
668 r_5, respectively. Results for the Aws pathway were also in line with predictions with
669 the majority of the mutations arising in the negative regulator AwsX or the interacting
670 part of AwsR (both decreasing r_3), with rare mutations in AwsO and other parts of
671 AwsR. For MwsR, loss-of-function mutations were expected primarily in the
672 phosphodiesterase (PDE) domain, but mutations were found in both diguanylate
673 cyclase and PDE domains. A structural analysis revealed that mutations were
674 clustered in the interface between the domains and unlikely to disrupt PDE function.
675 These mutations most likely change the dynamics between the two MwsR domains in
676 a way that increases production of c-di-GMP.

677

678 **Differences of mutational spectra with and without selection**

679 The null models – and especially null Model IV – successfully predicted the
680 mutational targets and explained mutation rates to WS when knowledge of mutational
681 hot spots was included. To what degree can such knowledge allow prediction of the
682 outcome of the original experimental evolution under selection (McDonald, et al.
683 2009)? A comparison between the frequencies of mutations isolated here (without
684 selection) with those isolated under selection, reveals several notable differences
685 (Figure 8). The most obvious difference is in use of the Wsp pathway, which is most
686 commonly used (15/24) under selection and yet produces WS types at a lower rate

687 than the Aws pathway. Differences are also apparent in the spectrum of *wsp*
688 mutations, with no *wspA* mutations being found under selection despite being the
689 most commonly mutated gene without selection (15/46), and the previous failure to
690 detect *wspR* mutants in a screen of 53 WS mutants (Goymer, et al. 2006).



691

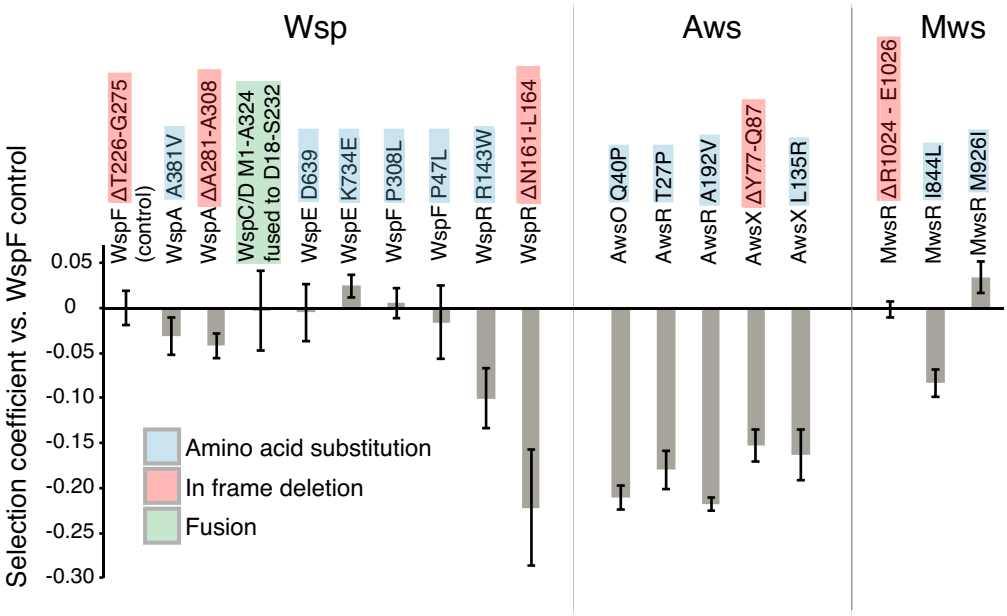
692 **Figure 8. WS mutations isolated with and without selection. (A)** The mutational
693 spectrum observed under selection suggests underrepresentation of WspA and WspR
694 compared to WspF and WspE. (**B**) Similar fitness effects of different Aws mutants
695 lead to similar patterns regardless of selective conditions. However the frequency of
696 all Aws mutants isolated under selection is much lower than expected from its high
697 mutation rate. (**C**) The low number of MwsR mutants isolated does not allow an
698 analysis of relative rates. Only within operon comparisons are valid for this figure as
699 the mutants isolated without selection had double deletions of the other operons.
700 Between operon mutation rates are available in Figure 2.

701

702 **Fitness of WS types**

703 The most obvious explanation for the differences in mutational spectra between WS
704 isolated with or without selection (Figure 8) is that certain mutants have a lower

705 fitness and thus their relative frequencies will be lower in the original evolution
706 experiment (McDonald, et al. 2009). We measured the fitness of representative WS
707 types with mutations in each of the mutated genes (*wspA*, *wspC/D*, *wspE*, *wspF*,
708 *wspR*, *awsX*, *awsR*, *awsO*, *mwsR*) in 1:1 competitions against a reference WspF
709 ΔT226-G275 deletion mutant marked with GFP (Figure 9). This type of fitness data
710 should be interpreted with caution because the fitness of WS mutants are frequency-
711 dependent (Rainey and Travisano 1998) and some WS mutants are superior in early
712 phase attachment as opposed to growth at the air-liquid interface (Lind, et al. 2015).
713 Nevertheless, the competition experiments provide an estimate of fitness when several
714 different WS mutants compete at the air-liquid interface (a likely situation given a
715 ~10^{-8} mutation rate to WS and a final population size of >10^{10}). The fitness data
716 account for the over- or under-representation of some WS mutants when grown under
717 selection (McDonald, et al. 2009) compared to those uncovered without selection (as
718 reported here).



719

720 **Figure 9. Fitness of different WS mutants.** Competitive fitness against a WspF
721 ΔT226-G275 reference strain was measured for representative mutations in the Wsp,

722 Aws, Mws pathways. Pairwise competitions were performed in quadruplicates and
723 error bars represent +/- one standard deviation. Full data, including statistical tests is
724 available in Figure 9 –source data.
725
726 The three *wspF* mutants, the *wspC-wspD* fusion, and the *wspE* mutants have similar
727 fitness ($P > 0.38$, except for WspE K734E that has slightly higher fitness ($P = 0.027$),
728 two-tailed *t*-tests). In contrast, both *wspA* mutants are slightly less fit ($P < 0.0214$,
729 two-tailed *t*-test) and both *wspR* mutants are severely impaired ($P < 0.00007$, two-
730 tailed *t*-test) (Figure 9). This sits in accord with previous work in which mutations
731 generating WS obtained with selection have been detected in *wspF* and *wspE*, but not
732 *wspA* or *wspR* (Goymer, et al. 2006; McDonald, et al. 2009). All *awsXRO* mutants
733 have similar lower fitness ($P < 10^{-6}$, two-tailed *t*-test) compared to the *wspF* reference
734 strain (Figure 9), which explain why under selection these are found at lower
735 frequencies compared to mutations in the *wsp* pathway (McDonald, et al. 2009)
736 despite a roughly two-fold higher mutation rate to WS.
737

738 **Discussion**

739 The issue of evolutionary predictability and the relative importance of stochastic
740 events compared to deterministic processes have a long history in evolutionary
741 biology (Darwin 1872; Simpson 1949; Jacob 1977; Gould 1989; Conway Morris
742 2003; Orgogozo 2015). Recent interest has been sparked by an increasing number of
743 observations that both phenotypic and molecular evolution, under certain
744 circumstances, can be remarkably repeatable (Colosimo, et al. 2005; Shindo, et al.
745 2005; Jost, et al. 2008; Barrick, et al. 2009; Lee and Marx 2012; Meyer, et al. 2012;
746 Zhen, et al. 2012; Herron and Doebeli 2013), but whether these cases are

representative for evolutionary processes in general remains to be determined. A related question, with greater potential for practical applications, is whether it is possible to forecast short-term evolutionary events and if so, then the challenge is to stipulate the data necessary to make successful predictions.

Our uniquely detailed knowledge of the WS experimental evolution system has provided a rare opportunity to determine the contributions of mutational bias and genetic architecture to the generation of new adaptive phenotypes, and consequently explore the limits of evolutionary forecasting. A thorough understanding of the function of the molecular species and their interactions allowed development of null models (see especially Model IV) that capture essential features of the genotype-to-phenotype map sufficient to predict the relative likelihood that evolution will follow each of the three principle pathways, along with specific mutational (genetic) targets. Even the most sophisticated of these models failed to forecast outcomes that matched the experimental data shown in Figure 2, however, the reason became apparent upon characterisation of the set of WS mutants obtained without selection: the presence of mutational hotspots. Armed with knowledge of sources of mutational bias it was a simple matter to refine Model IV leading to predictions matching those observed by experiment.

Problematic at the current time is inability to *a priori* detect all causes of mutational bias, however, it is likely that this will improve as understanding of the biochemical causes of bias improves and algorithms trained to recognise and detect nucleotide patterns indicative of bias are implemented. One specific deletion (ΔY77-Q87) in *awsX* was found to account for nearly half (20/41) of the mutations in the Aws

772 pathway. Thus, despite the existence of hundreds of possible mutations leading to WS
773 (this work and (McDonald, et al. 2009; McDonald, et al. 2011; Lind, et al. 2015)) one
774 single mutation accounts for more than one quarter of all WS mutations. While the six
775 base pair direct repeat flanking the deletion provides a convincing explanation for its
776 increased rate, it is not clear why this deletion would be ten times more common than
777 the ΔP34-A46 deletion in the same gene that is flanked by ten base pair repeats and
778 contains five base pairs identical to those from the ΔY77-Q87 deletion (Figure 5 –
779 source data 1). There are also instances where single base pair substitutions are
780 overrepresented: the AwsR T27P mutation is found in nine cases, while eight other
781 single pair substitutions in Aws were found only once. Consider further the fact that
782 *wspE* (a gene of ~2.3 kb), where changes to only four specific amino acids repeatedly
783 cause WS, and *wspF* (a gene of ~1 kb) where any mutation that disrupts function
784 results in WS (Figure 6A) contribute equally to the rates of at which new WS types
785 arise.
786
787 It is evident from these findings and from related studies (Pollock and Larkin 2004)
788 that there is need for detailed experimental measurement of local mutation rates in
789 specific systems. Such investigations stand to contribute to understanding of the
790 causes of mutational bias and the extent to which biases might be conserved among
791 related or even unrelated organisms. If local nucleotide sequence is the major
792 determinant, an estimate of mutation rate will apply strictly to very closely related
793 species, but if the dynamics of molecular processes, such as transcription and
794 replication (Sankar, et al. 2016), are major influences then estimates might be
795 applicable to a wider range of species.
796

797 Forecasting mutational routes to new phenotypes is one component of a
798 comprehensive forecasting strategy. The second requirement is ability to *a priori*
799 predict fitness effects. Solving this problem is the Holy Grail for predicting
800 evolution, but at the current time this is not possible. That it matters is made clear by
801 the observation that a subset of all possible WS mutants was found after experimental
802 evolution (Figure 8) due to inability of some to successfully compete with high fitness
803 WS types (Figure 9).
804
805 Direct measurement of the fitness effects of large numbers of mutations is difficult,
806 time-consuming, typically only possible for microbial species and against the spirit of
807 identifying *a priori* predictors of fitness effects. Future success likely rests on ability
808 to infer fitness from parameters such as estimated effects of mutations on
809 thermodynamic stability (Capriotti, et al. 2005; Dean and Thornton 2007; Rodrigues,
810 et al. 2016), molecular networks, evolutionary conservation of amino acid residues
811 (Glaser, et al. 2003; Ng and Henikoff 2003; Landau, et al. 2005; Kumar, et al. 2009;
812 Celniker, et al. 2013; Choi and Chan 2015) or machine-learning methods combining
813 several types of information (Bromberg and Rost 2007; Li, et al. 2009; Capriotti, et al.
814 2013; Yates, et al. 2014; Hecht, et al. 2015). Possibilities likely also exist to
815 extrapolate findings from a small number of mutations that are either directly
816 constructed and assayed in the laboratory or through fitness estimates of
817 polymorphism data from natural populations. Recent work on the prediction of the
818 fitness effects of random mutations in several genes suggests that in many cases large
819 effect mutations can be predicted using methods based on evolutionary conservation
820 (Lind, et al. 2017a).
821

822 On a less ambitious scale it may be possible to take advantage of the fact that the
823 distribution of fitness effects associated with mutations in single genes is often
824 bimodal – a consequence of many mutations causing complete loss-of-function rather
825 than intermediate deleterious effects (Sanjuan 2010; Hietpas, et al. 2011; Jacquier, et
826 al. 2013; Firnberg, et al. 2014; Lind, et al. 2016; Sarkisyan, et al. 2016; Lundin, et al.
827 2017). As such it maybe sufficient to know gene function plus consequences arising
828 from loss of function mutations and thus treat mutations within single genes as having
829 equivalent fitness (Sanjuan 2010; Jacquier, et al. 2013; Sarkisyan, et al. 2016; Lind, et
830 al. 2017a). Interestingly WS mutations occurring in the same gene typically show
831 similar fitness effects (Figure 9). While still requiring experimental data, this
832 simplification may fuel exploration of the relationship between large numbers of
833 mutations and their fitness effects and thus understanding of the extent to which
834 fitness effects are transferable between strains with different genetic backgrounds or
835 closely related species (Ungerer, et al. 2003; Pearson, et al. 2012; Wang, et al. 2014).
836
837 Taken at face value, our findings give reason to question the value of aspiring to
838 forecast evolution from first principles based on mechanistic understanding. But we
839 argue against such pessimism and point firstly to the value that stems from a clear
840 understanding of current limitations on forecasting: defining what is known, what is
841 not known, and what needs to be known. That it is possible to take knowledge of the
842 genotype-to-phenotype map and forecast with good accuracy targets and rates is an
843 important advance. There is good reason to suspect that the principles outlined here
844 and previously (Lind et al 2015) are transferable to other systems and even
845 generalizable (Lind 2018).
846

847 Evolutionary forecasting is in its infancy. In the short term it is likely to be most
848 successful for biological systems where there are experimental data on a large number
849 of independent evolutionary events, such as is the case for influenza, HIV and cancer
850 (Kouyos, et al. 2012; Fraser, et al. 2014; Lawrence, et al. 2014; Łuksza and Lässig
851 2014; Neher, et al. 2014; Eirew, et al. 2015). Evolution might appear idiosyncratic
852 indicating that every specific system requires detailed investigation, but our hope is
853 that deeper knowledge of the genotype-to-phenotype map, distribution of fitness
854 effects and mutational biases will allow short term forecasts to be produced using
855 modelling without the need for large-scale experimental studies. A major boost to
856 further refinement of evolutionary forecasting is likely to come from combining
857 coarse (top down) and fine-grained (bottom-up) approaches. Our demonstration that
858 simple null models of functional networks can produce quantitative predictions is a
859 step forward allowing predictions to be directly tested in other experimental systems
860 (Lind 2018).
861

862 # Materials and methods

Key Resources Table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
gene ()	*wspA*; PFLU1219	NA		
gene ()	*wspB*; PFLU1220	NA		
gene ()	*wspC*; PFLU1221	NA		
gene ()	*wspD*; PFLU1222	NA		
gene ()	*wspE*; PFLU1223	NA		
gene ()	*wspF*; PFLU1224	NA		
gene ()	*wspR*; PFLU1225	NA		
gene ()	*awsX*; PFLU5211, *yfiR*	NA		
gene ()	*awsR*; PFLU5210, *yfiN*, *tpbB*	NA		

gene ()	*awsO*; PFLU5209, *yfiB*	NA		
gene ()	*mwsR*; PFLU5329, *morA*	NA		
gene ()	wss; PFLU0300-PFL0309	NA		
strain, strain background ()	*Pseudomonas fluorescens* SBW25	PMID: 19432983; PMID: 8830243		
genetic reagent ()	PBR721; Δaws Δmws	PMID: 19704015		
genetic reagent ()	PBR713; Δwsp Δmws	PMID: 19704015		
genetic reagent ()	PBR712; Δwsp Δaws	PMID: 19704015		
genetic reagent ()	WT with pMSC	PMID: 17377582		
genetic reagent ()	PBR721 with pMSC	This work		Wsp pathway with pMSC
genetic reagent ()	PBR713 with pMSC	This work		Aws pathway with pMSC
genetic reagent ()	PBR712 with pMSC	This work		Mws pathway with pMSC
cell line ()				
transfected construct ()				
biological sample ()				
antibody				
recombinant DNA reagent	pMSC	PMID: 17377582		
recombinant DNA reagent	mini-Tn7(Gm)$P_{rrnB\ P1}$ gfp-a	PMID: 15186351		
sequence-based reagent	PFLU1219_1294R	Sigma-Aldrich		agacactgcggactggatct
sequence-based reagent	PFLU1219_UF	Sigma-Aldrich		ttgcctgtggagcgttctat
sequence-based reagent	PFLU1219_1130F	Sigma-Aldrich		accaggtggtggtgaccat
sequence-based reagent	PFLU1220_DR	Sigma-Aldrich		gacgatcactgcttcgatca
sequence-based reagent	wspD1r	Sigma-Aldrich		gttgcgttccagggcgag
sequence-based reagent	PFLU1221_DR	Sigma-Aldrich		acttatcgccatggataccg
sequence-based reagent	PFLU1221_UF	Sigma-Aldrich		tggatcggtgtaaacgacct
sequence-based reagent	PFLU1223_1511F	Sigma-Aldrich		atgacatcgtgcaactggaa
sequence-based reagent	PFLU1223_DR	Sigma-Aldrich		caggtccatcaggatcaggt
sequence-based reagent	PFLU1223_841F	Sigma-Aldrich		cgcccaagtgttgtatgaca
sequence-based reagent	PFLU1223_DR	Sigma-Aldrich		caggtccatcaggatcaggt
sequence-based reagent	PFLU1224_160R	Sigma-Aldrich		tccatcaccggcatgatca
sequence-based reagent	PFLU1224_DR	Sigma-Aldrich		ttttcgtcggtcttgatgtc
sequence-based reagent	PFLU1224_UF	Sigma-Aldrich		aagccagtttccacgatgac
sequence-based reagent	PFLU1225_DR	Sigma-Aldrich		gtgagggtacaccgagccta

sequence-based reagent	PFLU1225_UF	Sigma-Aldrich		tgccaaatgaacatgacctc
sequence-based reagent	PFLU5329_1888F	Sigma-Aldrich		cgtggattgggtgttcaaga
sequence-based reagent	PFLU5329_3118R	Sigma-Aldrich		cggttggtaatagagcacga
sequence-based reagent	wspF-seqR	Sigma-Aldrich		ataacggtcggcttgatacg
sequence-based reagent	Aws11f	Sigma-Aldrich		gctggttcagcttgatcgaaccc
sequence-based reagent	Aws16r	Sigma-Aldrich		atggatgctgccgatggttc
sequence-based reagent	Aws7f	Sigma-Aldrich		gcaaactcaccgcggatgag
sequence-based reagent	Aor	Sigma-Aldrich		cccgtcgattctccgtgcgc
peptide, recombinant protein				
commercial assay or kit				
chemical compound, drug				
software, algorithm	Ma-Sandri-Sarkar MLE estimator	PMID: 19369502		www.keshavsingh.org/protocols/FALCOR.html
software, algorithm	Julia v1.0	https://julialang.org		
software, algorithm	Matlab 9.4			
software, algorithm	Phyre2	PMID: 25950237		http://www.sbg.bio.ic.ac.uk/phyre2
software, algorithm	Geneious 7.1	https://www.geneious.com		
software, algorithm	FACSDiva 7.0	BD Biosciences		
other				

Strains and media

The strains used in the study are all *Pseudomonas fluorescens* SBW25 (Silby, et al. 2009) or derivatives thereof. The reporter construct (pMSC), used for isolation of WS mutants before selection, fused the P_{wss} promoter to a kanamycin resistance marker (*nptII*) (Fukami, et al. 2007; McDonald, et al. 2011). *P. fluorescens* strains with deletions of the *wsp* (PFLU1219-1225), *aws* (PFLU5209-5211) and *mws* (PFLU5329) operons were previously constructed as described by McDonald *et al.* (McDonald, et al. 2011). All experiments used King's medium B (KB) (King, et al. 1954), solidified

872 with 1.5% agar and incubation was at 28°C. All strains were stored in glycerol saline
873 at -80°C.
874
875 **Fluctuation tests and isolation of WS mutants before selection**
876 Strains with the pMSC reporter construct and either wild type genetic background or
877 double deletions of *aws*/*mws*, *wsp*/*mws* or *wsp*/*aws* were used to estimate mutation
878 rates to WS before selection. Overnight cultures were diluted to approximately 10^3
879 cfu/ml and 60 independent 110 µl cultures were grown for 16-19 h (OD600= 0.9-1.0)
880 with shaking (200 rpm) in 96-well plates before plating on KB plates with 30 mg/l
881 kanamycin. Viable counts were estimated by serial dilution and plating on KB agar.
882 One randomly chosen colony per independent culture with WS colony morphology
883 was restreaked once on KB agar. The assay was repeated at least four times for the
884 double deletion mutants and twice for the wild type strain in order to obtain enough
885 mutants to allow estimation of mutation rates. Mutations rates and confidence
886 intervals were estimated using the Ma-Sandri-Sarkar Maximum Likelihood Estimator
887 (Hall, et al. 2009) available at www.keshavsingh.org/protocols/FALCOR.html. The
888 mutation rates between the different strains were statistically evaluated using a t-test
889 as previously described (Rosche and Foster 2000), but this method has only been
890 shown to be valid in cases where total population size is not significantly different for
891 the strains used. In our assay this was not the case, as determined by ANOVA, and
892 therefore the results of the statistical analysis should be interpreted with caution. As
893 the estimated number of mutants per well was < 0.5 for all strains, the biasing effect
894 of differences in fitness between WS mutants is minimal.
895
896 **Sequencing**

897 Mutations causing the WS phenotype were identified by Sanger sequencing of
898 candidate genes in the remaining common pathway to WS, for example the *wsp*
899 operon for the *aws*/*mws* deletion strain. In a few cases where no mutations were
900 identified in the previously established WS target genes, we used genome sequencing
901 (Illumina HiSeq, performed by Macrogen Korea).
902
903 **Fitness assays**
904 Competition assays were performed as previously described (Lind, et al. 2015) by
905 mixing the WS mutant 1:1 with a reference strain labelled with green fluorescent
906 protein and measuring the ratio of each strain before and after static growth for 24 h
907 using flow cytometry (BD FACS Canto). We used a WspF ΔT226-G275 deletion
908 mutant as the reference strain because WspF mutants are the most commonly found
909 WS type when grown under selective conditions (McDonald, et al. 2009) and the in
910 frame deletion of 50 amino acids most likely represents a complete loss-of-function
911 mutation with minimal polar effects on the downstream *wspR*. Selection coefficients
912 per generation were calculated as $s = [\ln(R(t)/R(0))]/[t]$, as previously described
913 (Dykhuizen 1990) where R is the ratio of alternative WS mutant to WspF ΔT226-
914 G275 GFP and t the number of generations. Viable counts on KB plates of initial and
915 final populations were performed to calculate the number of generations. Stability of
916 colony morphologies were confirmed and data from microcosms with >5% smooth
917 colonies were excluded (two cases). Control competition experiments with isogenic
918 WspF ΔT226-G275 reference strains with and without GFP were used to correct for
919 the cost of the GFP marker. Control competitions were also used to determine the cost
920 of the double deletions and the reporter construct relative to a wild type genetic
921 background, for example an AwsX ΔY77-Q87 mutant in *Δwsp*/*Δmws* background

922 with pMSC was competed with a GFP labeled AwsX ΔY77-Q87 mutant in wild type
923 background. Competitions were performed in independently inoculated
924 quadruplicates for each strain with the number of replicates based on previous work
925 (Lind, et al. 2015).
926
927 **Homology models**
928 Homology models of the structure of WspA, WspE, WspR, AwsR, AwsO and MwsR
929 were made using Phyre2 in intensive mode (http://www.sbg.bio.ic.ac.uk/phyre2)
930 (Kelley, et al. 2015).
931
932 **Probability estimation in the null Model IV**
933 The differential equation models describe the interactions between proteins in each of
934 the three WS pathways. In order to solve the differential equations, two pieces of
935 information are required: i) the initial concentrations of the molecular species and ii)
936 the reaction rates. Although this information is unavailable a random-sampling
937 approach was used to generate different random sets of initial concentrations and
938 reaction rates. Each random set was used to establish a baseline of potential WS
939 expression making it possible to evaluate whether a set of mutations results in a WS
940 type. Effectively, this approach allows sampling of the probability distribution $P(WS$
941 $|m_i \in Wsp)$ used in our Bayesian model.
942
943 We randomly sample 1,000 different sets of reaction rates and initial concentrations
944 from uniform priors: reaction rates were sampled randomly from a uniform
945 distribution on log space (i.e. $10^{U[-2,2]}$) and initial concentrations of reactants were
946 sampled from a uniform distribution U[0,10]. For each set, the appropriate differential

947 equation model was integrated and the steady state concentration of the compounds
948 that correspond to a wrinkly spreader (RR in Aws, R* in Wsp and D* for Mws)
949 computed. This served as a baseline for the non-WS phenotype that was used for
950 comparison to determine whether combinations of mutations result in increased WS
951 expression. After obtaining the baseline, we implemented particular combinations of
952 enabling/disabling mutations ($a\ m_i$). Ideally, a distribution linking enabling/disabling
953 mutations to a fold change in reaction rates would be used, but this information is
954 unavailable. In order to progress the effect sizes for enabling and disabling mutations
955 were sampled from $10^{U[0,2]}$ and $10^{U[-2,0]}$, respectively, and then multiplied by the
956 reaction rates. The differential equations were then solved for the same time that it
957 took the baseline to reach steady state. The final concentration of R* (Figure 4A), RR
958 (Figure 4B) and D* (Figure 4C) was then compared to the baseline and the number of
959 times out of 1,000 that the WS-inducing compound increased served as an estimate of
960 $P\ (WS|m_i \in Wsp)$. The probability distribution stabilized by 500 random samples and
961 additional sampling did not produce significant changes (data not shown).
962
963 The absence of empirical data on reaction rates, initial concentrations, and expected
964 mutation effect size meant using a random sampling approach requiring estimates for
965 parameter ranges. Parameter ranges were chosen to be broad enough to capture
966 differences spanning several orders of magnitudes while allowing numerical
967 computations for solving the differential equations. To assess the effect of these
968 ranges on the results, the sampling procedure was repeated for WSP for three
969 different parameter regimes i) an expanded range for initial concentrations [0-50], ii)
970 an expanded range for reaction rates $10^{[-3,3]}$, iii) a compressed range for mutational

971 effect size $10^{[-1,1]}$. This analysis shows that qualitative results are robust to these
972 changes (see Figure S1).
973
974 **Supplementary information**
975 Figure 6 – source data. Table of all WS mutations in Wsp, Aws and Mws.
976 Figure 9 – source data. Data from fitness assay.
977 Supplementary text: Analysis of mutations and molecular effects. Source code, Figure
978 S1
979

980 ## Acknowledgements
981 This work was supported by Marsden Fund Council from New Zealand Government
982 funding, administered by the Royal Society of New Zealand.
983

984 ## Competing interests
985 The authors declare no competing interests.
986

987 ## Author contributions
988 **Peter A. Lind** Conceptualization, Methodology, Investigation, Data Curation,
989 Writing—original draft, Writing—review and editing, Visualization
990
991 **Eric Libby** Conceptualization, Methodology, Software, Formal analysis,
992 Investigation, Writing—original draft, Writing—review and editing, Visualization
993
994 **Jenny Herzog** Methodology, Investigation, Data Curation
995

996 **Paul B. Rainey** Conceptualization, Methodology, Validation, Writing—original
997 draft, Writing—review and editing, Visualization, Supervision, Project
998 administration, Funding acquisition
999

1000 ## References
1001

1002 Alberch P. 1991. From genes to phenotype: dynamical systems and evolvability.
1003 Genetica 84:5-11.

1004 Bailey SF, Blanquart F, Bataillon T, Kassen R. 2017. What drives parallel evolution?:
1005 How population size and mutational variation contribute to repeated evolution.
1006 Bioessays 39:1-9. http://dx.doi.org/10.1002/bies.201600176

1007 Bantinaki E, Kassen R, Knight CG, Robinson Z, Spiers AJ, Rainey PB. 2007.
1008 Adaptive divergence in experimental populations of Pseudomonas fluorescens. III.
1009 Mutational origins of wrinkly spreader diversity. Genetics 176:441-453.
1010 http://dx.doi.org/10.1534/genetics.106.069906

1011 Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF.
1012 2009. Genome evolution and adaptation in a long-term experiment with Escherichia
1013 coli. Nature 461:1243-1247. http://dx.doi.org/10.1038/nature08480

1014 Bromberg Y, Rost B. 2007. SNAP: predict effect of non-synonymous polymorphisms
1015 on function. Nucleic Acids Res 35:3823-3835. http://dx.doi.org/10.1093/nar/gkm238

1016 Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. 2013. WS-
1017 SNPs&GO: a web server for predicting the deleterious effect of human protein
1018 variants using functional annotation. BMC Genomics 14 Suppl 3:S6.
1019 http://dx.doi.org/10.1186/1471-2164-14-S3-S6

1020 Capriotti E, Fariselli P, Casadio R. 2005. I-Mutant2.0: predicting stability changes
1021 upon mutation from the protein sequence or structure. Nucleic Acids Res 33:W306-
1022 310. http://dx.doi.org/10.1093/nar/gki375

1023 Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T, Ben-
1024 Tal N. 2013. ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about
1025 Protein Function. Israel Journal of Chemistry 53:199-206.
1026 http://dx.doi.org/10.1002/ijch.201200096

1027 Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect
1028 of amino acid substitutions and indels. Bioinformatics 31:2745-2747.
1029 http://dx.doi.org/10.1093/bioinformatics/btv195

1030 Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Jr., Dickson M, Grimwood
1031 J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel
1032 evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science
1033 307:1928-1933. http://dx.doi.org/10.1126/science.1107239

1034 Conway Morris S. 2003. Life's solution inevitable humans in a lonely universe. New
1035 York u.a.: Cambridge Univ. Press.

1036 Darwin C. 1872. The origin of species by means of natural selection, or, The
1037 preservation of favoured races in the struggle for life. London: John Murray.

1038 De N, Navarro MV, Raghavan RV, Sondermann H. 2009. Determinants for the
1039 activation and autoinhibition of the diguanylate cyclase response regulator WspR. J
1040 Mol Biol 393:619-633. http://dx.doi.org/10.1016/j.jmb.2009.08.030

1041 de Visser JA, Krug J. 2014. Empirical fitness landscapes and the predictability of
1042 evolution. Nature reviews. Genetics 15:480-490. http://dx.doi.org/10.1038/nrg3744

1043 Dean AM, Thornton JW. 2007. Mechanistic approaches to the study of evolution: the
1044 functional synthesis. Nature reviews. Genetics 8:675-688.
1045 http://dx.doi.org/10.1038/nrg2160

1046 Dykhuizen DE. 1990. Experimental studies of natural selection in bacteria. Annu.
1047 Rev. Ecol. Syst. 21:373-398.

1048 Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, Gelmon K, Chia S, Mar C,
1049 Wan A, et al. 2015. Dynamics of genomic clones in breast cancer patient xenografts
1050 at single-cell resolution. Nature 518:422-426. http://dx.doi.org/10.1038/nature13952

1051 Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new
1052 mutations. Nature reviews. Genetics 8:610-618. http://dx.doi.org/10.1038/nrg2146

1053 Ferguson GC, Bertels F, Rainey PB. 2013. Adaptive Divergence in Experimental
1054 Populations of Pseudomonas fluorescens. V. Insight into the Niche Specialist "Fuzzy
1055 Spreader" Compels Revision of the Model Pseudomonas Radiation. Genetics.
1056 http://dx.doi.org/10.1534/genetics.113.154948

1057 Firnberg E, Labonte JW, Gray JJ, Ostermeier M. 2014. A comprehensive, high-
1058 resolution map of a gene's fitness landscape. Mol Biol Evol 31:1581-1592.
1059 http://dx.doi.org/10.1093/molbev/msu081

1060 Flowers JM, Hanzawa Y, Hall MC, Moore RC, Purugganan MD. 2009. Population
1061 genomics of the Arabidopsis thaliana flowering time gene network. Molecular
1062 biology and evolution 26:2475-2486. http://dx.doi.org/10.1093/molbev/msp161

1063 Foster PL, Lee H, Popodi E, Townes JP, Tang H. 2015. Determinants of spontaneous
1064 mutation in the bacterium Escherichia coli as revealed by whole-genome sequencing.
1065 Proceedings of the National Academy of Sciences of the United States of America
1066 112:E5990-5999. http://dx.doi.org/10.1073/pnas.1512136112

1067 Fraebel DT, Mickalide H, Schnitkey D, Merritt J, Kuhlman TE, Kuehn S. 2017.
1068 Environment determines evolutionary trajectory in a constrained phenotypic space.
1069 Elife 6. http://dx.doi.org/10.7554/eLife.24669

1070 Fraser C, Lythgoe K, Leventhal GE, Shirreff G, Hollingsworth TD, Alizon S,
1071 Bonhoeffer S. 2014. Virulence and pathogenesis of HIV-1 infection: an evolutionary
1072 perspective. Science 343:1243727. http://dx.doi.org/10.1126/science.1243727

1073 Fukami T, Beaumont HJ, Zhang XX, Rainey PB. 2007. Immigration history controls
1074 diversification in experimental adaptive radiation. Nature 446:436-439.
1075 http://dx.doi.org/10.1038/nature05629

1076 Galen SC, Natarajan C, Moriyama H, Weber RE, Fago A, Benham PM, Chavez AN,
1077 Cheviron ZA, Storz JF, Witt CC. 2015. Contribution of a mutational hot spot to
1078 hemoglobin adaptation in high-altitude Andean house wrens. Proceedings of the
1079 National Academy of Sciences of the United States of America 112:13958-13963.
1080 http://dx.doi.org/10.1073/pnas.1507300112

1081 Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N. 2003.
1082 ConSurf: identification of functional regions in proteins by surface-mapping of
1083 phylogenetic information. Bioinformatics 19:163-164.

1084 Gompel N, Prud'homme B. 2009. The causes of repeated genetic evolution. Dev Biol
1085 332:36-47. http://dx.doi.org/10.1016/j.ydbio.2009.04.040

1086 Gould SJ. 1989. Wonderful life : the Burgess Shale and the nature of history. New
1087 York: W.W. Norton.

1088 Goymer P, Kahn SG, Malone JG, Gehrig SM, Spiers AJ, Rainey PB. 2006. Adaptive
1089 divergence in experimental populations of Pseudomonas fluorescens. II. Role of the
1090 GGDEF regulator WspR in evolution and development of the wrinkly spreader
1091 phenotype. Genetics 173:515-526. http://dx.doi.org/10.1534/genetics.106.055863

1092 Hall BM, Ma CX, Liang P, Singh KK. 2009. Fluctuation analysis CalculatOR: a web
1093 tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis.
1094 Bioinformatics 25:1564-1565. http://dx.doi.org/10.1093/bioinformatics/btp253

1095 Hammerschmidt K, Rose CJ, Kerr B, Rainey PB. 2014. Life cycles, fitness
1096 decoupling and the evolution of multicellularity. Nature 515:75-79.
1097 http://dx.doi.org/10.1038/nature13884

1098 Hecht M, Bromberg Y, Rost B. 2015. Better prediction of functional effects for
1099 sequence variants. BMC Genomics 16 Suppl 8:S1. http://dx.doi.org/10.1186/1471-
1100 2164-16-S8-S1

1101 Heilbron K, Toll-Riera M, Kojadinovic M, MacLean RC. 2014. Fitness is strongly
1102 influenced by rare mutations of large effect in a microbial mutation accumulation
1103 experiment. Genetics 197:981-990. http://dx.doi.org/10.1534/genetics.114.163147

1104 Herron MD, Doebeli M. 2013. Parallel evolutionary dynamics of adaptive
1105 diversification in Escherichia coli. PLoS biology 11:e1001490.
1106 http://dx.doi.org/10.1371/journal.pbio.1001490

1107 Hietpas RT, Jensen JD, Bolon DN. 2011. Experimental illumination of a fitness
1108 landscape. Proc Natl Acad Sci U S A 108:7896-7901.
1109 http://dx.doi.org/10.1073/pnas.1016024108

1110 Huangyutitham V, Guvener ZT, Harwood CS. 2013. Subcellular clustering of the
1111 phosphorylated WspR response regulator protein stimulates its diguanylate cyclase
1112 activity. MBio 4:e00242-00213. http://dx.doi.org/10.1128/mBio.00242-13

1113 Jacob F. 1977. Evolution and tinkering. Science 196:1161-1166.

1114 Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, Bercot B, Petit
1115 E, Poulain J, Barnaud G, et al. 2013. Capturing the mutational landscape of the beta-
1116 lactamase TEM-1. Proceedings of the National Academy of Sciences of the United
1117 States of America 110:13067-13072. http://dx.doi.org/10.1073/pnas.1215206110

1118 Jost MC, Hillis DM, Lu Y, Kyle JW, Fozzard HA, Zakon HH. 2008. Toxin-resistant
1119 sodium channels: parallel adaptive evolution across a complete gene family.
1120 Molecular biology and evolution 25:1016-1024.
1121 http://dx.doi.org/10.1093/molbev/msn025

1122 Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web
1123 portal for protein modeling, prediction and analysis. Nat Protoc 10:845-858.
1124 http://dx.doi.org/10.1038/nprot.2015.053

1125 King EO, Ward MK, Raney DE. 1954. Two simple media for the demonstration of
1126 pyocyanin and fluorescin. The Journal of laboratory and clinical medicine 44:301-
1127 307.

1128 Kouyos RD, Leventhal GE, Hinkley T, Haddad M, Whitcomb JM, Petropoulos CJ,
1129 Bonhoeffer S. 2012. Exploring the complexity of the HIV-1 fitness landscape. PLoS
1130 genetics 8:e1002551. http://dx.doi.org/10.1371/journal.pgen.1002551

1131 Kram KE, Geiger C, Ismail WM, Lee H, Tang H, Foster PL, Finkel SE. 2017.
1132 Adaptation of Escherichia coli to Long-Term Serial Passage in Complex Medium:
1133 Evidence of Parallel Evolution. mSystems 2.
1134 http://dx.doi.org/10.1128/mSystems.00192-16

1135 Krasovec R, Richards H, Gifford DR, Hatcher C, Faulkner KJ, Belavkin RV,
1136 Channon A, Aston E, McBain AJ, Knight CG. 2017. Spontaneous mutation rate is a
1137 plastic trait associated with population density across domains of life. PLoS biology
1138 15:e2002731. http://dx.doi.org/10.1371/journal.pbio.2002731

1139 Kugelberg E, Kofoid E, Andersson DI, Lu Y, Mellor J, Roth FP, Roth JR. 2010. The
1140 tandem inversion duplication in Salmonella enterica: selection drives unstable
1141 precursors to final mutation types. Genetics 185:65-80.
1142 http://dx.doi.org/10.1534/genetics.110.114074

1143 Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous
1144 variants on protein function using the SIFT algorithm. Nat Protoc 4:1073-1081.
1145 http://dx.doi.org/10.1038/nprot.2009.86

1146 Laland KN, Uller T, Feldman MW, Sterelny K, Muller GB, Moczek A, Jablonka E,
1147 Odling-Smee J. 2015. The extended evolutionary synthesis: its structure, assumptions
1148 and predictions. Proceedings. Biological sciences / The Royal Society 282:20151019.
1149 http://dx.doi.org/10.1098/rspb.2015.1019

1150 Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. 2005.
1151 ConSurf 2005: the projection of evolutionary conservation scores of residues on
1152 protein structures. Nucleic Acids Res 33:W299-302.
1153 http://dx.doi.org/10.1093/nar/gki370

1154 Lässig M, Mustonen V, Walczak AM. 2017. Predicting evolution. Nature Ecology &
1155 Evolution 1:0077. http://dx.doi.org/10.1038/s41559-017-0077

1156 Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR,
1157 Meyerson M, Gabriel SB, Lander ES, Getz G. 2014. Discovery and saturation
1158 analysis of cancer genes across 21 tumour types. Nature 505:495-501.
1159 http://dx.doi.org/10.1038/nature12912

1160 Lee MC, Marx CJ. 2012. Repeated, selection-driven genome reduction of accessory
1161 genes in experimental populations. PLoS genetics 8:e1002651.
1162 http://dx.doi.org/10.1371/journal.pgen.1002651

1163 Levinson G, Gutman GA. 1987. Slipped-strand mispairing: a major mechanism for
1164 DNA sequence evolution. Molecular biology and evolution 4:203-221.
1165 http://dx.doi.org/10.1093/oxfordjournals.molbev.a040442

1166 Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD,
1167 Radivojac P. 2009. Automated inference of molecular mechanisms of disease from
1168 amino acid substitutions. Bioinformatics 25:2744-2750.
1169 http://dx.doi.org/10.1093/bioinformatics/btp528

1170 Lind PA. 2018. Evolutionary forecasting of phenotypic and genetic outcomes of
1171 experimental evolution in Pseudomonas. bioRxiv. http://dx.doi.org/10.1101/342261

1172 Lind PA, Andersson DI. 2008. Whole-genome mutational biases in bacteria.
1173 Proceedings of the National Academy of Sciences of the United States of America
1174 105:17878-17883. http://dx.doi.org/10.1073/pnas.0804445105

1175 Lind PA, Arvidsson L, Berg OG, Andersson DI. 2017a. Variation in Mutational
1176 Robustness between Different Proteins and the Predictability of Fitness Effects.
1177 Molecular biology and evolution 34:408-418.
1178 http://dx.doi.org/10.1093/molbev/msw239

1179 Lind PA, Arvidsson L, Berg OG, Andersson DI. 2016. Variation in Mutational
1180 Robustness between Different Proteins and the Predictability of Fitness Effects.
1181 Molecular biology and evolution. http://dx.doi.org/10.1093/molbev/msw239

1182 Lind PA, Farr AD, Rainey PB. 2017b. Evolutionary convergence in experimental
1183 Pseudomonas populations. ISME J 11:589-600.
1184 http://dx.doi.org/10.1038/ismej.2016.157

1185 Lind PA, Farr AD, Rainey PB. 2015. Experimental evolution reveals hidden diversity
1186 in evolutionary pathways. Elife 4. http://dx.doi.org/10.7554/eLife.07074

1187 Lovett ST. 2004. Encoded errors: mutations and rearrangements mediated by
1188 misalignment at repetitive DNA sequences. Mol Microbiol 52:1243-1253.
1189 http://dx.doi.org/10.1111/j.1365-2958.2004.04076.x

1190 Luksza M, Lassig M. 2014. A predictive fitness model for influenza. Nature 507:57-
1191 61. http://dx.doi.org/10.1038/nature13087

1192 Lundin E, Tang PC, Guy L, Nasvall J, Andersson DI. 2017. Experimental
1193 determination and prediction of the fitness effects of random point mutations in the

1194 biosynthetic enzyme HisA. Molecular biology and evolution.
1195 http://dx.doi.org/10.1093/molbev/msx325

1196 Luria SE, Delbruck M. 1943. Mutations of Bacteria from Virus Sensitivity to Virus
1197 Resistance. Genetics 28:491-511.

1198 Lynch M. 2010. Rate, molecular spectrum, and consequences of human mutation.
1199 Proceedings of the National Academy of Sciences of the United States of America
1200 107:961-968. http://dx.doi.org/10.1073/pnas.0912629107

1201 Maharjan RP, Ferenci T. 2017. A shifting mutational landscape in 6 nutritional states:
1202 Stress-induced mutagenesis as a series of distinct stress input-mutation output
1203 relationships. PLoS biology 15:e2001477.
1204 http://dx.doi.org/10.1371/journal.pbio.2001477

1205 Malone JG, Jaeger T, Manfredi P, Dotsch A, Blanka A, Bos R, Cornelis GR, Haussler
1206 S, Jenal U. 2012. The YfiBNR signal transduction mechanism reveals novel targets
1207 for the evolution of persistent Pseudomonas aeruginosa in cystic fibrosis airways.
1208 PLoS Pathog 8:e1002760. http://dx.doi.org/10.1371/journal.ppat.1002760

1209 Malone JG, Jaeger T, Spangler C, Ritz D, Spang A, Arrieumerlou C, Kaever V,
1210 Landmann R, Jenal U. 2010. YfiBNR mediates cyclic di-GMP dependent small
1211 colony variant formation and persistence in Pseudomonas aeruginosa. PLoS Pathog
1212 6:e1000804. http://dx.doi.org/10.1371/journal.ppat.1000804

1213 Martin P, Sun L, Hood DW, Moxon ER. 2004. Involvement of genes of genome
1214 maintenance in the regulation of phase variation frequencies in Neisseria
1215 meningitidis. Microbiology 150:3001-3012. http://dx.doi.org/10.1099/mic.0.27182-0

McCandlish DM, Stoltzfus A. 2014. Modeling evolution using the probability of fixation: history and implications. Quarterly Review of Biology 89:225-252.

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

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

Meyer JR, Dobias DT, Weitz JS, Barrick JE, Quick RT, Lenski RE. 2012. Repeatability and contingency in the evolution of a key innovation in phage lambda. Science 335:428-432. http://dx.doi.org/10.1126/science.1214449

Moxon ER, Rainey PB, Nowak MA, Lenski RE. 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. Current biology : CB 4:24-33.

Neher RA, Russell CA, Shraiman BI. 2014. Predicting evolution from the shape of genealogical trees. eLife 3. http://dx.doi.org/10.7554/eLife.03568

Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31:3812-3814.

O'Connor JR, Kuwada NJ, Huangyutitham V, Wiggins PA, Harwood CS. 2012. Surface sensing and lateral subcellular localization of WspA, the receptor in a chemosensory-like system leading to c-di-GMP production. Mol Microbiol 86:720-729. http://dx.doi.org/10.1111/mmi.12013

1238 Orgogozo V. 2015. Replaying the tape of life in the twenty-first century. Interface
1239 Focus 5:20150057. http://dx.doi.org/10.1098/rsfs.2015.0057

1240 Orr HA. 2005. The genetic theory of adaptation: a brief history. Nature reviews.
1241 Genetics 6:119-127. http://dx.doi.org/10.1038/nrg1523

1242 Ostrowski EA, Woods RJ, Lenski RE. 2008. The genetic basis of parallel and
1243 divergent phenotypic responses in evolving populations of Escherichia coli.
1244 Proceedings. Biological sciences / The Royal Society 275:277-284.
1245 http://dx.doi.org/10.1098/rspb.2007.1244

1246 Parkinson JS. 2010. Signaling mechanisms of HAMP domains in chemoreceptors and
1247 sensor kinases. Annu Rev Microbiol 64:101-122.
1248 http://dx.doi.org/10.1146/annurev.micro.112408.134215

1249 Pearson VM, Miller CR, Rokyta DR. 2012. The consistency of beneficial fitness
1250 effects of mutations across diverse genetic backgrounds. Plos One 7:e43864.
1251 http://dx.doi.org/10.1371/journal.pone.0043864

1252 Phippen CW, Mikolajek H, Schlaefli HG, Keevil CW, Webb JS, Tews I. 2014.
1253 Formation and dimerization of the phosphodiesterase active site of the Pseudomonas
1254 aeruginosa MorA, a bi-functional c-di-GMP regulator. FEBS Lett 588:4631-4636.
1255 http://dx.doi.org/10.1016/j.febslet.2014.11.002

1256 Pigliucci M. 2010. Genotype-phenotype mapping and the end of the 'genes as
1257 blueprint' metaphor. Philosophical transactions of the Royal Society of London.
1258 Series B, Biological sciences 365:557-566. http://dx.doi.org/10.1098/rstb.2009.0241

1259 Pollock DD, Larkin JC. 2004. Estimating the degree of saturation in mutant screens.
1260 Genetics 168:489-502. http://dx.doi.org/10.1534/genetics.103.024430

1261 Rainey PB, Remigi P, Farr AD, Lind PA. 2017. Darwin was right: where now for
1262 experimental evolution? Curr Opin Genet Dev 47:102-109.
1263 http://dx.doi.org/10.1016/j.gde.2017.09.003

1264 Rainey PB, Travisano M. 1998. Adaptive radiation in a heterogeneous environment.
1265 Nature 394:69-72. http://dx.doi.org/10.1038/27900

1266 Reams AB, Roth JR. 2015. Mechanisms of gene duplication and amplification. Cold
1267 Spring Harb Perspect Biol 7:a016592. http://dx.doi.org/10.1101/cshperspect.a016592

1268 Reijns MA, Kemp H, Ding J, de Proce SM, Jackson AP, Taylor MS. 2015. Lagging-
1269 strand replication shapes the mutational landscape of the genome. Nature 518:502-
1270 506. http://dx.doi.org/10.1038/nature14183

1271 Rice MS, Dahlquist FW. 1991. Sites of deamidation and methylation in Tsr, a
1272 bacterial chemotaxis sensory transducer. J Biol Chem 266:9746-9753.

1273 Richardson AR, Stojiljkovic I. 2001. Mismatch repair and the regulation of phase
1274 variation in Neisseria meningitidis. Mol Microbiol 40:645-655.

1275 Rodrigues JV, Bershtein S, Li A, Lozovsky ER, Hartl DL, Shakhnovich EI. 2016.
1276 Biophysical principles predict fitness landscapes of drug resistance. Proceedings of
1277 the National Academy of Sciences of the United States of America 113:E1470-1478.
1278 http://dx.doi.org/10.1073/pnas.1601441113

1279 Rosche WA, Foster PL. 2000. Determining mutation rates in bacterial populations.
1280 Methods 20:4-17. http://dx.doi.org/10.1006/meth.1999.0901

1281 Roth JR, Benson N, Galitsky T, Haack K, Lawrence JG, Miesel L. 1996.
1282 Rearrangements of the Bacterial Chromosome: Formation and
1283 Applications. In: Neidhardt FC, editor. Escherichia coli and Salmonella: Cellular and
1284 Molecular Biology. Washington, D.C.: 2256-2276. p. 2256-2276.

1285 Sanjuan R. 2010. Mutational fitness effects in RNA and single-stranded DNA viruses:
1286 common patterns revealed by site-directed mutagenesis studies. Philos Trans R Soc
1287 Lond B Biol Sci 365:1975-1982. http://dx.doi.org/10.1098/rstb.2010.0063

1288 Sankar TS, Wastuwidyaningtyas BD, Dong Y, Lewis SA, Wang JD. 2016. The nature
1289 of mutations induced by replication-transcription collisions. Nature.
1290 http://dx.doi.org/10.1038/nature18316

1291 Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV,
1292 Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, et al. 2016. Local fitness
1293 landscape of the green fluorescent protein. Nature 533:397-401.
1294 http://dx.doi.org/10.1038/nature17995

1295 Seier T, Padgett DR, Zilberberg G, Sutera VA, Jr., Toha N, Lovett ST. 2011. Insights
1296 into mutagenesis using Escherichia coli chromosomal lacZ strains that enable
1297 detection of a wide spectrum of mutational events. Genetics 188:247-262.
1298 http://dx.doi.org/10.1534/genetics.111.127746

1299 Shewaramani S, Finn TJ, Leahy SC, Kassen R, Rainey PB, Moon CD. 2017.
1300 Anaerobically Grown Escherichia coli Has an Enhanced Mutation Rate and Distinct
1301 Mutational Spectra. PLoS genetics 13:e1006570.
1302 http://dx.doi.org/10.1371/journal.pgen.1006570

1303 Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C. 2005.
1304 Role of FRIGIDA and FLOWERING LOCUS C in determining variation in
1305 flowering time of Arabidopsis. Plant physiology 138:1163-1173.
1306 http://dx.doi.org/10.1104/pp.105.061309

1307 Silby MW, Cerdeno-Tarraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston
1308 GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SA, et al. 2009. Genomic and
1309 genetic analyses of diversity and plant interactions of Pseudomonas fluorescens.
1310 Genome biology 10:R51. http://dx.doi.org/10.1186/gb-2009-10-5-r51

1311 Simpson GG. 1949. The meaning of evolution : a study of the history of life and of its
1312 significance for man. New Haven: Yale University Press.

1313 Spiers AJ, Bohannon J, Gehrig SM, Rainey PB. 2003. Biofilm formation at the air-
1314 liquid interface by the Pseudomonas fluorescens SBW25 wrinkly spreader requires an
1315 acetylated form of cellulose. Molecular microbiology 50:15-27.

1316 Spiers AJ, Kahn SG, Bohannon J, Travisano M, Rainey PB. 2002. Adaptive
1317 divergence in experimental populations of Pseudomonas fluorescens. I. Genetic and
1318 phenotypic bases of wrinkly spreader fitness. Genetics 161:33-46.

1319 Steinrueck M, Guet CC. 2017. Complex chromosomal neighborhood effects
1320 determine the adaptive potential of a gene under selection. Elife 6.
1321 http://dx.doi.org/10.7554/eLife.25100

1322 Stern DL. 2013. The genetic causes of convergent evolution. Nature reviews.
1323 Genetics 14:751-764. http://dx.doi.org/10.1038/nrg3483

1324 Stern DL, Orgogozo V. 2009. Is genetic evolution predictable? Science 323:746-751.
1325 http://dx.doi.org/10.1126/science.1158997

1326 Stoltzfus A, McCandlish DM. 2017. Mutational Biases Influence Parallel Adaptation.
1327 Molecular biology and evolution 34:2163-2172.
1328 http://dx.doi.org/10.1093/molbev/msx180

1329 Tautz D, Renz M. 1984. Simple sequences are ubiquitous repetitive components of
1330 eukaryotic genomes. Nucleic Acids Res 12:4127-4138.

1331 Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD,
1332 Gaut BS. 2012. The molecular diversity of adaptive convergence. Science 335:457-
1333 461. http://dx.doi.org/10.1126/science.1212986

1334 Ungerer MC, Linder CR, Rieseberg LH. 2003. Effects of genetic background on
1335 response to selection in experimental populations of Arabidopsis thaliana. Genetics
1336 163:277-286.

1337 Viswanathan M, Lacirignola JJ, Hurley RL, Lovett ST. 2000. A novel mutational
1338 hotspot in a natural quasipalindrome in Escherichia coli. J Mol Biol 302:553-564.
1339 http://dx.doi.org/10.1006/jmbi.2000.4088

1340 Wang AD, Sharp NP, Agrawal AF. 2014. Sensitivity of the distribution of mutational
1341 fitness effects to environment, genetic background, and adaptedness: a case study
1342 with Drosophila. Evolution 68:840-853. http://dx.doi.org/10.1111/evo.12309

1343 Weinreich DM, Delaney NF, Depristo MA, Hartl DL. 2006. Darwinian evolution can
1344 follow only very few mutational paths to fitter proteins. Science 312:111-114.
1345 http://dx.doi.org/10.1126/science.1123539

1346 Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE. 2006. Tests of parallel
1347 molecular evolution in a long-term experiment with Escherichia coli. Proceedings of
1348 the National Academy of Sciences of the United States of America 103:9107-9112.
1349 http://dx.doi.org/10.1073/pnas.0602917103

1350 Xu M, Yang X, Yang XA, Zhou L, Liu TZ, Fan Z, Jiang T. 2016. Structural insights
1351 into the regulatory mechanism of the Pseudomonas aeruginosa YfiBNR system.
1352 Protein Cell 7:403-416. http://dx.doi.org/10.1007/s13238-016-0264-7

1353 Yampolsky LY, Stoltzfus A. 2001. Bias in the introduction of variation as an
1354 orienting factor in evolution. Evol Dev 3:73-83.

1355 Yates CM, Filippis I, Kelley LA, Sternberg MJ. 2014. SuSPect: enhanced prediction
1356 of single amino acid variant (SAV) phenotype using network features. J Mol Biol
1357 426:2692-2701. http://dx.doi.org/10.1016/j.jmb.2014.04.026

1358 Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. 2012. Parallel
1359 molecular evolution in an herbivore community. Science 337:1634-1637.
1360 http://dx.doi.org/10.1126/science.1226630
1361