Predicting mutational routes to new adaptive phenotypes

Peter A. Lind\textsuperscript{1,2}, Eric Libby\textsuperscript{1,3,4}, Jenny Herzog\textsuperscript{1} and Paul B. Rainey\textsuperscript{1,5,6}

\textsuperscript{1}New Zealand Institute for Advanced Study, Massey University at Albany, Auckland, 0745, New Zealand.
\textsuperscript{2}Department of Molecular Biology, Umeå University, SE-901 87 Umeå, Sweden.
\textsuperscript{3}Santa Fe Institute, Santa Fe, New Mexico, United States of America.
\textsuperscript{4}Department of Mathematics, Umeå University, SE-901 87 Umeå, Sweden.
\textsuperscript{5}Department of Microbial Population Biology, Max Planck Institute for Evolutionary Biology, Plön 24306, Germany.
\textsuperscript{6}Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI ParisTech), 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
Thus, the functions of gene products and their regulatory interactions provide information on likely mutational targets underpinning particular phenotypes. For example, consider a hypothetical structural determinant subject to both positive and negative regulation and whose over-expression generates a given adaptive phenotype. Assuming a uniform distribution of mutational events, mutations in the negative regulator (and not the positive activator) will be the primary cause of the adaptive phenotype. This follows from the fact that loss-of-function mutations are more common than gain-of-function mutations. Indeed, an emerging rule indicates that phenotypes determined by genetic pathways that are themselves subject to negative regulation are most likely to arise by loss-of-function mutations in negative regulatory components (McDonald, et al. 2009; Tenaillon, et al. 2012; Lind, et al. 2015; Fraebel, et al. 2017).

Mutation is not equally likely at each nucleotide of a given genome (Lind and Andersson 2008; Lynch 2010; Seier, et al. 2011; Foster, et al. 2015; Reijns, et al. 2015; Sankar, et al. 2016; Stoltzfus and McCandlish 2017). Numerous instances of mutational bias have been reported. Prime examples are simple sequence repeats such as homopolymeric nucleotide tracts or di-, tri- and tetrameric repeats that mutate at high frequency via slipped strand mispairing (Levinson and Gutman 1987). These readily identifiable sequences define contingency loci in obligate human pathogens and commensals (Moxon, et al. 1994) and are widespread in eukaryotic genomes (Tautz and Renz 1984). The behaviour of contingency loci can be further modulated by defects in components of methyl-directed mismatch repair systems (Richardson and Stojilkovic 2001; Martin, et al. 2004; Hammerschmidt, et al. 2014; Heilbron, et
Certain palindromic structures also lead to mutational bias (Viswanathan, et al. 2000; Lovett 2004) and promote amplification events that can increase target size for additional mutations (Roth, et al. 1996; Kugelberg, et al. 2010; Reams and Roth 2015) Transition-transversion bias (Stoltzfus and McCandlish 2017) and elevated mutation rates at CpG sites (Galen, et al. 2015) can further skew the distributions of mutational effects. Additional bias arises from the chromosomal neighbourhood of genes under selection (Steinrueck and Guet 2017), the location of genes with regard to interactions with DNA replication/transcription machineries (Sankar, et al. 2016), and environmental factors that affect not only mutation rate but also the spectra of mutational events (Krasovec, et al. 2017; Maharjan and Ferenci 2017; Shewaramani, et al. 2017).

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

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

**Figure 1.** The *Pseudomonas fluorescens* SBW25 wrinkly spreader model. (A) Selection for access to oxygen allows wrinkly spreader (WS) mutants to invade the ancestral smooth (SM) population in static microcosms. WS mutants form a mat at the air-liquid interface through increased expression of the main structural
component, cellulose, encoded by the \textit{wss} operon. Expression of cellulose is controlled by the second messenger c-di-GMP, which is produced by diguanylate cyclases (DGCs). Mutations in the \textit{wsp}, \textit{aws} and \textit{mws} operons, that activate their respective DGCs (WspR, AwsR, MwsR), are the primary mutational pathways to WS.

(B) When a reporter construct connecting expression of the \textit{wss} operon to resistance to kanamycin is used under shaken non-selective conditions, WS mutants can be isolated without the biasing influence of natural selection. This allows estimation of the mutation rate to WS and an unbiased spectrum of mutations defining the mutational target. Fitness can then be assayed in competition with a common reference strain.

McDonald \textit{et al}. (McDonald, \textit{et al}. 2009) showed that each time the tape of WS evolution is re-run mutations generating the adaptive type arise in one of three DGC-encoding pathways (Wsp, Aws, or Mws) (Figure 1A). Subsequent work revealed that when these three pathways are eliminated from the ancestral type evolution proceeds along multiple new pathways (Lind, \textit{et al}. 2015) resulting in WS types whose fitnesses are indistinguishable from those arising via mutations in Wsp, Aws, or Mws. Parallel evolution of WS involving preferential usage of the Wsp, Aws and Mws pathways is therefore not explained by selection: repeated use of Wsp, Aws and Mws stems from the fact that these pathways are subject to negative regulation and thus, relative to pathways subject to positive regulation, or requiring promoter-activating mutations, gene fusion events, or other rare mutations, present a large mutational target (Lind, \textit{et al}. 2015).
Given repeatability of WS evolution, knowledge of the Wsp / Aws / Mws pathways, plus genetic tools for mechanistic investigation — including capacity to obtain WS mutants in the absence of selection — the WS system offers a rare opportunity to explore the extent to which knowledge of the genotype-to-phenotype map can lead mechanistic models for evolutionary forecasting. Our findings show that short-term mechanistic-level predictions of mutational pathways are possible, but also draw attention to challenges that stem from current inability to a priori predict locus-specific mutational biases and environment-specific fitness effects.

Results

Obtaining an unbiased measure of pathway-specific mutation rates to WS

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

In the ancestral SM genotype the cellulose promoter is inactive in shaken King’s Medium B (KB) broth (Spiers, et al. 2002) and thus the strain is sensitive to kanamycin. When a WS-causing mutation occurs, the wss promoter becomes active
resulting in a kanamycin-resistant WS type (Fukami, et al. 2007; McDonald, et al. 2011). Individual growth of this set of three genotypes in shaken KB, combined with plating to detect kanamycin-resistant mutants, makes possible a fluctuation assay (Luria and Delbruck 1943; Hall, et al. 2009) from which a direct measure of the rate at which WS mutants arise can be obtained. Importantly, because WS types are maladapted in shaken broth culture, the screen for kanamycin-resistant clones allows the pathway-specific mutation rate to WS to be obtained without the biasing effects of selection for growth at the air-liquid interface (Figure 1B). The results are shown in Figure 2.

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

The mutation rate was highest for the Aws pathway \( (6.5 \times 10^{-9}) \); approximately double that of Wsp \( (3.7 \times 10^{-9}) \) and an order of magnitude higher than that of the Mws pathway \( (0.74 \times 10^{-9}) \) (Figure 2). The rate at which WS mutants arose from the
ancestral genotype in which the three pathways are intact \((11.2 \times 10^{-9})\) was approximately the sum of the rates for the three pathways \((11.0 \times 10^{-9})\) confirming that the Wsp, Aws and Mws pathways are the primary routes by which WS types evolve (Lind, et al. 2015). That the Aws pathway has the greatest capacity to generate WS is surprising given the smaller target size (three genes and 2.3 kb compared to seven genes (8.4 kb) in the Wsp pathway).

**The genotype-to-phenotype map underpinning WS evolution**

Much is known about the function and interactions among components of each of the three focal pathways. This knowledge allows development of models that capture the dynamic nature of each pathway and thus allow predictions as to the likelihood that evolution will precede via each of the three mutational routes. An unresolved issue is the extent to which these models match experimental findings. Following a brief description of each pathway we describe the models.

The 8.4 kb Wsp pathway is a chemotaxis-like system (Goymer, et al. 2006; Guvener and Harwood 2007; Römling, et al. 2013; Micali and Endres 2016) composed of seven genes with the first six genes \((wspA-wspF)\) being transcribed as a single unit and the last \((wspR\) from its own promoter (Bantinaki, et al. 2007). WspA (PFLU1219) is a methyl-accepting chemotaxis (MCP) protein that forms a complex with the CheW-like scaffolding proteins WspB (PFLU1220) and WspD (PFLU1222). WspA senses environmental stimuli and transmits the information via conformational changes in the WspA/WspB/WspD complex to effect activity of WspE (PFLU1223), a CheA/Y hybrid histidine kinase response regulator. WspE activates both the WspR (PFLU1225) diguanylate cyclase (DGC) and the CheB-like methylesterase WspF
(PFLU1224) following transference of an active phosphoryl group. The activity of WspA is modulated by methylation by the constitutively active CheR-like methyltransferase WspC (PFLU1221) that transfers methyl groups to conserved glutamine residues on WspA. The demethylase WspF serves to remove these groups when in the phosphorylated active form. WS mutants are known to arise by mutations in the WspF negative regulator and also in the WspE kinase (McDonald, et al. 2009). In vitro manipulations of WspR that abolish repression of the DGC domain by the response regulator domain are known, but have never been observed to occur in experimental populations (Goymer, et al. 2006).

The 2.3 kb aws operon contains three genes transcribed from a single promoter (awsXRO). Homologous genes in Pseudomonas aeruginosa (yfiRNB, PA1121-1119) have been characterised in detail (Malone, et al. 2010; Malone, et al. 2012; Xu, et al. 2016). The outer membrane lipoprotein AwsO (PFLU5209) has an OmpA domain, a signal peptide and binds to peptidoglycan. AwsO is thought to be the sensor whose activity is modulated in response to envelope stress (Malone, et al. 2012). AwsO sequesters the periplasmic protein AwsX (PFLU5211) at the outer membrane. AwsX functions as a negative regulator of the DGC AwsR (PFLU5210) in the inner membrane. Both increased binding of AwsX to AwsO or loss of negative regulation by inactivation of the interaction between AwsX and AwsR can lead to WS (McDonald, et al. 2009; Malone, et al. 2010; Malone, et al. 2012).

The 3.9 kb mwsR gene (PFLU5329) is known as morA (PA4601) in Pseudomonas aeruginosa, and encodes a predicted membrane protein with both a DGC domain that produces c-di-GMP and a phosphodiesterase (PDE) domain that degrades c-di-GMP.
Little is known of the molecular details determining its function, but both catalytic domains appear to be active (Phippen, et al. 2014). Deletion of the PDE domain results in a WS phenotype with activity being dependent on a functional DGC domain (McDonald, et al. 2009).

**Development of mechanistic models predicting mutational routes to WS**

If the specific effects of changing each nucleotide (and sets of nucleotides) were known then models for each pathway would not be required. Here, we show how knowledge of genetic architecture can be used to build models that predict the likelihood that mutations generating WS types will arise in a given pathway – and even in specific genes. In the following section we present four null models that incorporate increasing levels of information concerning the genotype-to-phenotype map. The goal of these models is two-fold: firstly to demonstrate that incorporation of knowledge of genetic architecture allows development of models with explanatory value and secondly, to define minimal necessary information for reliable forecasting. The results are summarised in Figure 3, which displays the experimental data from 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 \((p_d)\), \(p_d = 0.0001, p_e = 0.001\).
Null Model I is intentionally naïve. It uses sequence length as a proxy for mutational target size, but ignores genetic organisation, function of predicted proteins and interaction among proteins. The model assumes that mutational target size is proportional to the number of nucleotides at a given locus and thus the probability that a given pathway is used to generate WS relative to another is simply the ratio of the probability of generating WS for the two focal pathways. For any one pathway the probability that a mutation generates WS is given by \(1-(1-p)^n\), where \(p\) is the probability of a mutation at a nucleotide and \(n\) is the number of nucleotides in the pathway. If the mutation probability is low such that the expected number of mutations in a pathway is below 1, that is, \(np<<1\) then the binomial approximation can be used: \(1-(1-p)^n = np\). Thus the probability that evolution follows the Wsp pathway over the Aws pathway is: \(8400 \frac{p_1}{2300} \frac{p_2}{p_2} = 3.65 \frac{p_1}{p_2}\), where \(p_1\) and \(p_2\) are the mutation rates for each pathway. Assuming equivalency of mutation rate, \(p_1 = p_2\), evolution is predicted to proceed via the Wsp pathway 3.65 times more often than via Aws, with evolution predicted to proceed via Mws 1.65 times more often than Aws.

Comparison with experimental data shows a departure both in terms of the priority of pathways used by evolution and the frequency of pathway usage (Figure 3A versus 3B).

Null Model II builds on Model I but only in a marginal sense. It recognises that nucleotides defining loci of interest are organised into genes, and therefore adopts gene number as a proxy for mutational target size. As above, the probability that a pathway is used is \(1-(1-p)^n = np\), but in this instance \(n\) is the number of genes. Model II predicts that mutations in Wsp generate WS types 2.33 times more often than
mutations in Aws, with mutations in Mws being 3-fold less likely to generate WS compared to mutations in Aws (Figure 3C). This marginal adjustment makes little difference to the fit between experimental data and predictions.

Past work has shown the explanatory value of information that comes from knowledge of gene (protein) function and interactions (McDonald, et al. 2009; Lind, et al. 2015). These form the basis of Null Model III. The relevant functions and interactions are depicted in Figure 4 as reaction diagrams that reflect how changes in different interactions affect the production of WS types. Organizing interactions within each pathway according to reactions has the advantage that it allows for a standard mathematical description of the biochemical dynamics using differential equations (see Figure 4 and Figure 4 - figure supplements 1 to 3). An additional advantage is that such an approach allows ready incorporation of new experimental data including interaction type and interaction strength.
Figure 4. Reaction diagrams for null models III and IV. Reaction diagrams show the interactions governing production of a wrinkly spreader in (A) Wsp, (B) Aws and (C) Mws pathways. The blue circles indicate molecular components, the red circles indicate reactions, and arrows indicate which molecular components are reactants and which are products. Full descriptions of the models including all reactions and the resulting systems of differential equations are found in Figure 4 figure supplement 1 for Wsp, Figure 4 figure supplement 2 for Aws and Figure 4 figure supplement 3 for Mws. The black edged circles are the reporter proteins, which is the activated forms of the diguanylate cyclases that directly determine whether a pathway generates a wrinkly spreader. Increased production of any reporter leads to increased c-di-GMP concentration causing increased cellulose production and thereby a wrinkly spreader.

**Figure 4 - figure supplement 1. Wsp model (A)** Description of functional interactions (B) Description of molecular reaction (C) Differential equations describing the dynamics of the Wsp pathway. The activity of WspA is modulated by methylation, where it is activated by the CheR-like methyltransferase WspC (PFLU1221). The CheB-like methylesterase WspF (PFLU1224) functions as a negative regulator. Modulation of WspR activity through changes in oligomerization 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.

**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.
Figure 4 - figure supplement 3. Mws model (A) Description of functional
interactions (B) Description of molecular reactions (C) Differential equations
describing the dynamics of the Aws pathway

One immediate consequence of this approach is that production of WS types is
described entirely through biochemical reactions. Knowing whether a WS-type is
generated by a particular mutation amounts to determining which reaction rates are
altered by the mutation and whether those changes affect the likelihood of producing
WS types. This leads to an approach where mutations in components can be classified
according to their effects on reaction rates: enabling mutations increase reaction rates,
whereas disabling mutations decrease reaction rates.

Based on the mathematical models for the pathways, each reaction rate can be altered
(either increased or decreased) to generate higher levels of the key WS effector (the
active DGC). For example, in the Wsp pathway, if reactions 2 or 6 experience a
disabling mutation, or alternatively, if any of the reactions 1, 3, 4, or 5 experience an
enabling mutation, then the level of activated (phosphorylated) WspR is increased.
For now, it is assumed that increased activity of the relevant DGC generates a WS
type. Reactions whose enabling / disabling mutations increase the amount of WS
effector are termed enabling / disabling reactions.

By comparing the number of disabling and enabling reactions in different pathways it
is possible to calculate the relative likelihood that evolution uses each mutational
pathway. The Wsp pathway encompasses two disabling and four enabling reactions,
whereas the Aws pathway is defined by one disabling and three enabling reactions.
Mws encompasses a single disabling reaction and two enabling reactions. If disabling mutations are as likely as enabling mutations, then the likelihood that a pathway will be used by evolution is simply the ratio of the total number of enabling and disabling reactions: 6:4:3 (Wsp:Aws:Mws). If instead, disabling mutations are much more likely than enabling mutations, then enabling reactions can be ignored and the likelihood that a pathway will be used by evolution becomes simply the ratio of the number of disabling reactions: 2:1:1 (Wsp:Aws:Mws) (Figure 3D left hand panel). If the reverse is true, then the likelihood of use is a ratio of the number of enabling reactions: 4:3:2 (Wsp:Aws:Mws) (Figure 3D right hand panel). In all cases, mutations in the Wsp pathway are predicted to be 1.3-2 times more likely to generate WS than mutations in the Aws pathway, with the Aws pathway being the target of evolution 1-1.5 times more often than the Mws pathway. Although still inflating the importance of the Wsp pathway relative to Aws, Mws is more prominent than in null Model II.

Null Model IV takes a further step toward mechanistic accuracy by endowing interactions with pleiotropic and continuous effects. This removes two simplifying assumptions of null Model III likely to limit predictive power. In Model III mutations affect one reaction at a time, but mutations in certain components, for example wspA, affect more than a single reaction (WspA appears in reactions 1-4). Additionally, in Model III, changes in reaction rate are assumed to be binary, however reactions may have a range of effects on the evolution of WS types.

To accommodate pleiotropic effects, null Model IV systematically considers all combinations of enabling and disabling changes to reaction rates and determines the likelihood that a WS type is generated. An example of one set of the possible
mutations ($m_i$) in Wsp is $1, -1, 0, 0, 0, 0$ (an increase in $r_1$, a decrease in $r_2$, but no change in $r_3, r_4, r_5, r_6$ (Figure 3A)). Since the Wsp pathway has six reaction rates, this amounts to $3^6$ or 729 total combinations. However, note that reaction 3 does not share any reactants or products with reactions 5 or 6. Thus, mutations such as $0, 0, 1, 0, 1, 0, 0, 0, 1$ are not considered because they require mutations in two separate genetic components.

To accommodate a range of effects null Model IV simulates enabling / disabling changes of different magnitudes and determines the resulting effect on the respective effector DGC (see Materials and Methods). Briefly, the approach addresses the lack of information concerning biochemical reaction rates and molecular concentrations in the mathematical models describing WS-producing pathway dynamics. By repeatedly sampling from the space of all possible reaction rates, initial concentrations, and magnitudes of effects, this approach computes the probability that a particular set of mutations ($m_i$), e.g. $1, -1, 0, 0, 0, 0$, results in a wrinkly spreader. This probability is represented as the conditional probability $P(WS | m_i \in Wsp)$, which motivates a Bayesian formulation to compare the relative probability that the different pathways produce WS. To this end, the probability that a particular pathway will be used is decomposed into two terms: the probability that a particular set of mutations ($m_i$) occurs in Wsp (or Aws, or Mws) represented as $P(m_i \in Wsp)$ and the probability that those mutations give rise to a wrinkly spreader represented as $P(WS | m_i \in Wsp)$ (or Aws, or Mws).

$$P(WS \cap m \cap Wsp) = \sum_i P(WS|m_i \in Wsp)P(m_i \in Wsp)$$ (1)
To estimate $P (m_i \in Wsp)$ we assume fixed probabilities of enabling and disabling changes and compute the product. Thus, the probability of $m_i = 1, -1, 0, 0, 0, 0$ is

$$p_e p_d (1 - p_e - p_d)^4,$$

where $p_e$ is the probability of a mutation with an enabling effect and $p_d$ is the probability of a mutation with a disabling effect. Recognising the value of accommodating the possibility of localised mutational bias we note that $p_e$ and $p_d$ can be adjusted for the affected reactants. The second term, $P (WS | m_i \in Wsp)$, relies on our sampling methodology and describes the probability that a set of disabling/enabling changes of different magnitudes will yield a WS type (see Materials and Methods).

Despite the mechanistic advances incorporated into null Model IV the Wsp pathway is still predicted to be the pathway most commonly used by evolution. The extreme cases in which disabling mutations are more probable than enabling mutations (and vice versa) are shown in Figure 3E. The results of simulations for the full range of $p_d$ and $p_e$ values are shown in Figure 5. Figure 5A shows that the Wsp pathway is predicted to be the target of mutation 1.5 - 1.9 times more often than the Aws pathway while Figure 5B shows that the Mws pathway is predicted to be the target of mutation 0.4 - 0.5 times less often than the Aws pathway. While these results agree with the experimental data showing Mws to be least likely pathway to be followed, the predictions are at odds with the data showing WS types to be twice as likely to arise from mutation in Aws, versus Wsp. The causes of this discrepancy are accounted for in the next section “analysis of mutants reveals sources of mutational bias”.

22
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
The rate of disabling mutations is much higher than the rate of enabling mutations. The majority of WS mutations in Wsp are produced by changes to $r_2$ and $r_6$.

**Figure 5 - figure supplement 2. Relative contributions of reaction rates for Aws 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 rate of disabling mutations is much higher than the rate of enabling mutations the majority of WS mutations in Wsp are produced by changes to $r_3$.

The Bayesian approach of null Model IV has additional utility: it allows prediction of genes likely to be affected by mutation. Predictions as to specific genetic targets come from appraisal of the relative importance of each reaction in terms of generating WS types (Figure 5 – supplement 1 and 2). While it is recognised that a majority of WS mutations arise from defects in negative regulators of DGCs, such as WspF and AwsX (McDonald, et al. 2009; Lind, et al. 2015), further predictions are possible.
based on impacts of alterations in gene function on reaction rates. For example, with reference to the Wsp pathway (Figure 5 – figure supplement 1), there are two reactions (2 and 6) that affect WspF function: $r_2$ describes the rate of removal of methyl groups from the signalling complex and $r_6$ the rate at which WspF is activated by transfer of active phosphoryl groups from the WspE kinase. Loss-of-function (disabling) mutations being much more common than gain-of-function (enabling) mutations means that both WspF and WspE are likely targets. The null model predicts that in the area of parameter space in which Wsp is most likely compared to Aws, 45% of the time WS will be generated when the second reaction, $r_2$, is altered (Figure 5 – figure supplement 1). The same is true for reaction $r_6$. Thus the presence of a negative regulator is predicted to extend the mutational target size well beyond the gene itself. This is also true for Aws, where $r_3$ is the main contributor to the WS phenotype in the case where disabling change is more common than enabling change. Here mutations are predicted not only in the negative regulator AwsX, but also in the interacting region of the DGC AwsR (Figure 5 – figure supplement 2).

Loss-of-function mutations in negative regulators and their interacting partners are not the only predicted targets. For Wsp $r_1, r_3, r_4,$ and $r_5$ are altered approximately 5% of the time in the parameter region where disabling mutations are more common than enabling mutations, but contribute more when the rate of enabling mutations is increased (Figure 5 – figure supplement 1). Enabling mutations based on null Model IV are likely to be found in WspC (increasing $r_1$), WspABD (increasing $r_3$), WspABD/WspE (increasing $r_4$) and WspR (increasing $r_5$) (Figure 4A). For Aws, enabling mutations are predicted to increase reaction $r_1$ by mutations causing constitutive activation of AwsO, $r_2$ increasing binding of AwsO and AwsX and $r_4$
increasing formation of the active AwsR dimer (Figure 4B, Figure 5 – figure supplement 2).

In summary, high rates of WS mutations are predicted for wspF, wspE, wspA, awsX and awsR with lower rates for wspC, wspR and awsO. Several of these predictions sit in accord with previous experimental observations, however, notable are predictions that evolution might also target wspA and wspR, two genes that have not previously been identified as mutational causes of WS types (McDonald, et al. 2009).

Analysis of mutants reveals sources of mutational bias

There are several reasons why predictions from the models might be out of kilter with experimental data on mutation rates. We firstly looked to the distribution of WS generating mutations among the 109 mutants collected during the course of the fluctuation assays. Of the 109 mutants, 105 harboured a mutation in wsp (46 mutants), aws (41 mutants) or mws (18 mutants) (Figure 6A, Figure 6 – source data 1). The remaining four had mutations in previously described rare pathways (PFLU0085, PFLU0183), again confirming that these non-focal pathways produce just a fraction of the total set of WS mutants (Lind, et al. 2015).
Figure 6. Mutational targets. (A) 105 independent mutations in the wsp (n=46), aws (n=41) and mws (n=18) operons were identified. Numbers of independent mutants are shown in brackets. Full details on the mutations are available in Figure 6 – source data 1. (B) Diversity of mutations with area proportional to mutation rate (Figure 2). Two mutations (AwsX ΔY77-Q87 and AwsR T27P) contribute 41% of all mutations to WS suggesting that these are mutational hot spots. The distribution of mutations for each of the three pathways is indicative of bias. As shown in Figure 6B, almost 29% of all WS-causing mutations (adjusted for differences in mutation rates between the three pathways) were due to an identical 33
base pair in-frame deletion in *awsX* (Δt229-g261, ΔY77-Q87), while a further 13 %
were due to an identical mutation (79 a>c, T27P) in *awsR*. At least 41 different
mutations in Aws lead to WS: if mutation rates were equal for these sites the
probability of observing 20 identical mutations would be extremely small. In fact 10
million random samplings from the observed distribution of mutations failed to
recover this bias. While the Wsp pathway also contains sites that were mutated more
than once (six positions were mutated twice, one site three times and one five times),
sources of mutational bias in Wsp were less evident than in Aws (Figure 6B).

**Incorporating mutational bias into null models**

The mathematical models presented above assumed no mutational bias, but the null
models can be modified to incorporate such bias. With focus on the AwsX hotspot,
we show this for models II-IV.

Null Model II: The probability of a mutation at the hotspot is *kp* where *k>*1 and *p* is
the probability of a mutation in non-hotspot locations. The probability that a pathway
is used is: *kp + (1-kp)(1-(1-p)^{(n-1)})* where *kp* is the probability of a mutation at the
hotspot and *(1-kp)(1-(1-p)^{(n-1)})* is the probability there is no mutation at the hotspot,
but that there is a mutation elsewhere in the pathway. Using the binomial
approximation, the probability a given pathway is used is: *kp + (1-kp)(n-1)p*, which
for Aws is *kp+(1-kp)2p*. Figure 7A shows that if the mutation rate at the hotspot
increases 10-fold (*k = 10.3*) then Aws is used 1.8 times more often than Wsp. While
this matches experimental data, the same 10-fold increase applied to Mws results in
mismatch. Only with a reduced mutation rate at the same locus (*k – 6.8*) do predicted
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 \( p_d = 0.001, p_e = 0.001 \) 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
dependent on the rates of enabling and enabling mutations. When the rate of disabling mutations is much higher than the rate of enabling mutations the majority of WS mutations in Wsp are produced by changes to $r_3$.

Null Model III: The mutational hotspot in the Aws pathway occurs in AwsX, which is involved in both an enabling and a disabling reaction. If the hotspot increases the rate of a disabling reaction by a factor $k$ where $k>1$, then the probability the pathway is used is:

$$k p_d + (1-k p_d) (1-(1-p_d)^{a_d-1}) + (1-k p_d) (1-p_d)^{a_d-1} (1-(1-p_e)^n),$$

which is the sum of probabilities of the events the hotspot is used, the hotspot is not used but a disabling change occurs, and the hotspot is not used but only an enabling change occurs. Figure 7C shows that if $k$ is between 4 and 5.3, then there is a good fit to experimental data. The actual value depends on the probabilities of enabling/disabling change and which reaction has the hotspot. For Mws the mutational hotspot must have a factor between 10 and 15 (Figure 7D).

Null Model IV: The mutational hotspot in $awsX$ affects reactions rates $r_2$ and $r_3$ in the Aws system of differential equations (Figure 4B, Figure 4 – figure supplement 2). If the mutation factor is between 3.3-5.6 for the Wsp / Aws comparison (Figure 7E) or 4.8-7.6 for the Mws / Aws comparison (Figure 7F), then predictions match observed data. Thus Model IV is the only one where the same five-fold change in the probability of enabling / disabling change predicts observed data for all three mutational pathways (Figure 7G). This model, updated with hotspot data, predicts
that the Aws pathway is more likely to generate WS types than Wsp for all probability
values (see Figure 7H). If disabling changes are more probable than enabling changes,
then \( r_3 \) dominates the predicted rate, whereas when enabling changes are more
common \( r_2 \) is the major contributor (Figure 7 – figure supplement 1). Knowledge of
the functional interaction between AwsX and AwsR allows interpretation of the likely
effect of the hotspot mutations in AwsX and AwsR as disabling change in \( r_3 \) (Table
1).

**Analysis of mutants reveals mutational targets and effects**

Null Model IV of the Wsp, Aws, and Mws pathways allow estimation of the relative
probability that a WS is produced by enabling or disabling changes in a certain
reaction rate. This means that in order to evaluate the success of our predictions, the
105 WS mutations must be connected to their likely effects on reaction rates. For
some mutations this is straightforward. If a mutation completely disables the negative
regulator WspF, this will reduce \( r_2 \) and \( r_6 \), thereby producing a wrinkly spreader. In
other cases interpretations of likely effects are more difficult and requires knowledge
of the molecular functions of the proteins and their interactions, which needs to be
obtained from literature, bioinformatics and structural modelling. For example,
mutations in WspA can potentially produce WS by disruption of the interaction with
WspF (reducing \( r_2 \)), increase methylation by WspC (increasing \( r_1 \)), increase signalling
rate (\( r_3 \)) or increase phosphorylation of WspE (\( r_4 \)). A summary of the interpretation of
the molecular effects of mutations and how they are connected to reaction rates is
available in Table 1 with a more detailed description in the supplementary text.

**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)

  - 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)
| Amino acid substitutions in between signal peptide and OmpA domain | Constitutive activation and sequestering of AwsX | Increase $r_2$ |
|---------------------------------------------------------------|-----------------------------------------------|--------------|
| AwsO OmpA domain                                              |                                               |              |

| 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 $r_1$ |
|---------------------------------------------------------------|-----------------------------------------------|--------------|
| MwsR EAL and DGC domains                                      |                                               |              |

| Amino acid substitutions near DGC active site | Loss of feedback regulation or changes to interdomain interaction | Decrease $r_2$, increase $r_1$ |
|---------------------------------------------------------------|-----------------------------------------------|--------------|
| MwsR DGC active site                                         |                                               |              |
Mutations were found in five genes in the Wsp pathway. The majority of the mutations were found in the negative regulator WspF or its interacting partners WspE and WspA. These mutations likely reduced reaction rates of $r_2$ and $r_6$ in accord with predictions of null Model IV. The remaining mutations in WspC and WspR were less common, in accord with predictions and likely due to increases in reaction rates $r_1$ and $r_5$, respectively. Results for the Aws pathway were also in line with predictions with the majority of the mutations arising in the negative regulator AwsX or the interacting part of AwsR (both decreasing $r_3$), with rare mutations in AwsO and other parts of AwsR. For MwsR, loss-of-function mutations were expected primarily in the phosphodiesterase (PDE) domain, but mutations were found in both diguanylate cyclase and PDE domains. A structural analysis revealed that mutations were clustered in the interface between the domains and unlikely to disrupt PDE function. These mutations most likely change the dynamics between the two MwsR domains in a way that increases production of c-di-GMP.

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

The null models – and especially null Model IV – successfully predicted the mutational targets and explained mutation rates to WS when knowledge of mutational hot spots was included. To what degree can such knowledge allow prediction of the outcome of the original experimental evolution under selection (McDonald, et al. 2009)? A comparison between the frequencies of mutations isolated here (without selection) with those isolated under selection, reveals several notable differences (Figure 8). The most obvious difference is in use of the Wsp pathway, which is most commonly used (15/24) under selection and yet produces WS types at a lower rate
than the Aws pathway. Differences are also apparent in the spectrum of \(\textit{wsp}\) mutations, with no \(\textit{wspA}\) mutations being found under selection despite being the most commonly mutated gene without selection (15/46), and the previous failure to detect \(\textit{wspR}\) mutants in a screen of 53 WS mutants (Goymer, et al. 2006).

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

**Fitness of WS types**

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

Figure 9. Fitness of different WS mutants. Competitive fitness against a WspF ΔT226-G275 reference strain was measured for representative mutations in the Wsp,
Aws, Mws pathways. Pairwise competitions were performed in quadruplicates and error bars represent +/- one standard deviation. Full data, including statistical tests is available in Figure 9 – source data.

The three wspF mutants, the wspC-wspD fusion, and the wspE mutants have similar fitness ($P > 0.38$, except for WspE K734E that has slightly higher fitness ($P = 0.027$), two-tailed t-tests). In contrast, both wspA mutants are slightly less fit ($P < 0.0214$, two-tailed t-test) and both wspR mutants are severely impaired ($P < 0.00007$, two-tailed t-test) (Figure 9). This sits in accord with previous work in which mutations generating WS obtained with selection have been detected in wspF and wspE, but not wspA or wspR (Goymer, et al. 2006; McDonald, et al. 2009). All awsXRO mutants have similar lower fitness ($P < 10^{-6}$, two-tailed t-test) compared to the wspF reference strain (Figure 9), which explain why under selection these are found at lower frequencies compared to mutations in the wsp pathway (McDonald, et al. 2009) despite a roughly two-fold higher mutation rate to WS.

**Discussion**

The issue of evolutionary predictability and the relative importance of stochastic events compared to deterministic processes have a long history in evolutionary biology (Darwin 1872; Simpson 1949; Jacob 1977; Gould 1989; Conway Morris 2003; Orgogozo 2015). Recent interest has been sparked by an increasing number of observations that both phenotypic and molecular evolution, under certain circumstances, can be remarkably repeatable (Colosimo, et al. 2005; Shindo, et al. 2005; Jost, et al. 2008; Barrick, et al. 2009; Lee and Marx 2012; Meyer, et al. 2012; Zhen, et al. 2012; Herron and Doebele 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
pathway. Thus, despite the existence of hundreds of possible mutations leading to WS (this work and (McDonald, et al. 2009; McDonald, et al. 2011; Lind, et al. 2015)) one single mutation accounts for more than one quarter of all WS mutations. While the six base pair direct repeat flanking the deletion provides a convincing explanation for its increased rate, it is not clear why this deletion would be ten times more common than the ΔP34-A46 deletion in the same gene that is flanked by ten base pair repeats and contains five base pairs identical to those from the ΔY77-Q87 deletion (Figure 5 – source data 1). There are also instances where single base pair substitutions are overrepresented: the AwsR T27P mutation is found in nine cases, while eight other single pair substitutions in Aws were found only once. Consider further the fact that wspE (a gene of ~2.3 kb), where changes to only four specific amino acids repeatedly cause WS, and wspF (a gene of ~1 kb) where any mutation that disrupts function results in WS (Figure 6A) contribute equally to the rates of at which new WS types arise.

It is evident from these findings and from related studies (Pollock and Larkin 2004) that there is need for detailed experimental measurement of local mutation rates in specific systems. Such investigations stand to contribute to understanding of the causes of mutational bias and the extent to which biases might be conserved among related or even unrelated organisms. If local nucleotide sequence is the major determinant, an estimate of mutation rate will apply strictly to very closely related species, but if the dynamics of molecular processes, such as transcription and replication (Sankar, et al. 2016), are major influences then estimates might be applicable to a wider range of species.
Forecasting mutational routes to new phenotypes is one component of a comprehensive forecasting strategy. The second requirement is ability to *a priori* predict fitness effects. Solving this problem is the Holy Grail for predicting evolution, but at the current time this is not possible. That it matters is made clear by the observation that a subset of all possible WS mutants was found after experimental evolution (Figure 8) due to inability of some to successfully compete with high fitness WS types (Figure 9).

Direct measurement of the fitness effects of large numbers of mutations is difficult, time-consuming, typically only possible for microbial species and against the spirit of identifying *a priori* predictors of fitness effects. Future success likely rests on ability to infer fitness from parameters such as estimated effects of mutations on thermodynamic stability (Capriotti, et al. 2005; Dean and Thornton 2007; Rodrigues, et al. 2016), molecular networks, evolutionary conservation of amino acid residues (Glaser, et al. 2003; Ng and Henikoff 2003; Landau, et al. 2005; Kumar, et al. 2009; Celniker, et al. 2013; Choi and Chan 2015) or machine-learning methods combining several types of information (Bromberg and Rost 2007; Li, et al. 2009; Capriotti, et al. 2013; Yates, et al. 2014; Hecht, et al. 2015). Possibilities likely also exist to extrapolate findings from a small number of mutations that are either directly constructed and assayed in the laboratory or through fitness estimates of polymorphism data from natural populations. Recent work on the prediction of the fitness effects of random mutations in several genes suggests that in many cases large effect mutations can be predicted using methods based on evolutionary conservation (Lind, et al. 2017a).
On a less ambitious scale it may be possible to take advantage of the fact that the
distribution of fitness effects associated with mutations in single genes is often
bimodal – a consequence of many mutations causing complete loss-of-function rather
than intermediate deleterious effects (Sanjuan 2010; Hietpas, et al. 2011; Jacquier, et
al. 2013; Firnberg, et al. 2014; Lind, et al. 2016; Sarkisyan, et al. 2016; Lundin, et al.
2017). As such it maybe sufficient to know gene function plus consequences arising
from loss of function mutations and thus treat mutations within single genes as having
equivalent fitness (Sanjuan 2010; Jacquier, et al. 2013; Sarkisyan, et al. 2016; Lind, et
al. 2017a). Interestingly WS mutations occurring in the same gene typically show
similar fitness effects (Figure 9). While still requiring experimental data, this
simplification may fuel exploration of the relationship between large numbers of
mutations and their fitness effects and thus understanding of the extent to which
fitness effects are transferable between strains with different genetic backgrounds or
closely related species (Ungerer, et al. 2003; Pearson, et al. 2012; Wang, et al. 2014).

Taken at face value, our findings give reason to question the value of aspiring to
forecast evolution from first principles based on mechanistic understanding. But we
argue against such pessimism and point firstly to the value that stems from a clear
understanding of current limitations on forecasting: defining what is known, what is
not known, and what needs to be known. That it is possible to take knowledge of the
genotype-to-phenotype map and forecast with good accuracy targets and rates is an
important advance. There is good reason to suspect that the principles outlined here
and previously (Lind et al 2015) are transferable to other systems and even
generalizable (Lind 2018).
Evolutionary forecasting is in its infancy. In the short term it is likely to be most successful for biological systems where there are experimental data on a large number of independent evolutionary events, such as is the case for influenza, HIV and cancer (Kouyos, et al. 2012; Fraser, et al. 2014; Lawrence, et al. 2014; Łuksza and Lässig 2014; Neher, et al. 2014; Eirew, et al. 2015). Evolution might appear idiosyncratic indicating that every specific system requires detailed investigation, but our hope is that deeper knowledge of the genotype-to-phenotype map, distribution of fitness effects and mutational biases will allow short term forecasts to be produced using modelling without the need for large-scale experimental studies. A major boost to further refinement of evolutionary forecasting is likely to come from combining coarse (top down) and fine-grained (bottom-up) approaches. Our demonstration that simple null models of functional networks can produce quantitative predictions is a step forward allowing predictions to be directly tested in other experimental systems (Lind 2018).

Materials and methods

| 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, yflR | NA           |             |                        |
| gene ()                           | awsR; PFLU5210, yflN, tpbB | NA       |             |                        |
| gene () |  |  |
| --- | --- | --- |
| gene () | awso; PFLU5209, yfB | NA |
| gene () | mwsR; PFLU5329, morA | NA |
| gene () | wss; PFLU0300-PFLU0309 | 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)PrrnP1gfp-α | PMID: 15186351 |
| sequence-based reagent | PFLU1219_1294R Sigma-Aldrich | agacactgcggactggatct |
| sequence-based reagent | PFLU1219_1294R Sigma-Aldrich | ttgcctgagacggtctctat |
| sequence-based reagent | PFLU1219_1130F Sigma-Aldrich | accaggtgggtgtgacat |
| sequence-based reagent | PFLU1220_DR Sigma-Aldrich | gacatcactgtagatca |
| sequence-based reagent | wspD1r Sigma-Aldrich | gttgccctccaggccag |
| sequence-based reagent | PFLU1221_DR Sigma-Aldrich | actatgcacattgataccg |
| sequence-based reagent | PFLU1221_UF Sigma-Aldrich | tggatgtgattacccgct |
| sequence-based reagent | PFLU1223_1511F Sigma-Aldrich | atgacatcgtgcaactggaa |
| sequence-based reagent | PFLU1223_DR Sigma-Aldrich | caggtccatcaggtgaggt |
| sequence-based reagent | PFLU1223_841F Sigma-Aldrich | cgcccaagtgtgtagac |
| sequence-based reagent | PFLU1223_DR Sigma-Aldrich | caggtccatcaggtgaggt |
| sequence-based reagent | PFLU1224_160R Sigma-Aldrich | tcctacccggcatgataca |
| sequence-based reagent | PFLU1224_DR Sigma-Aldrich | tttgcctgatctgat |
| sequence-based reagent | PFLU1224_UF Sigma-Aldrich | aagccagttccagcatgac |
| sequence-based reagent | PFLU1225_DR Sigma-Aldrich | gtgaggggtacacgcagctg |
| Sequence-based reagent | PFLU1225_UF | Sigma-Aldrich | tgccaaatgaacatgacctc |
|------------------------|-------------|---------------|---------------------|
| Sequence-based reagent | PFLU5329_1888F | Sigma-Aldrich | cggtgatggtgtcaaga |
| Sequence-based reagent | PFLU5329_3118R | Sigma-Aldrich | cggfagcatagcgccga |
| Sequence-based reagent | wspF-seqR | Sigma-Aldrich | ataacggtggtgtcag |
| Sequence-based reagent | Aws11f | Sigma-Aldrich | gctgggtcagcttgatgccc |
| Sequence-based reagent | Aws16r | Sigma-Aldrich | atgatgctgcatggttcc |
| Sequence-based reagent | Aws7f | Sigma-Aldrich | gcaacctcaccgcggatagag |
| Sequence-based reagent | Aor | Sigma-Aldrich | ccgctgatttctcggcgc |
| **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 |  |  |
| **Software, algorithm** | Matlab 9.4 |  |  |
| **Software, algorithm** | Phyre2 | PMID: 25950237 | http://www.sbg.bio.ic.ac.uk/phyre2 |
| **Software, algorithm** | Geneious 7.1 |  |  |
| **Software, algorithm** | FACSDiva 7.0 |  |  |
| **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$_{wsp}$ 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
with 1.5% agar and incubation was at 28°C. All strains were stored in glycerol saline at -80°C.

Fluctuation tests and isolation of WS mutants before selection

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

Sequencing
Mutations causing the WS phenotype were identified by Sanger sequencing of candidate genes in the remaining common pathway to WS, for example the wsp operon for the aws/mws deletion strain. In a few cases where no mutations were identified in the previously established WS target genes, we used genome sequencing (Illumina HiSeq, performed by Macrogen Korea).

**Fitness assays**

Competition assays were performed as previously described (Lind, et al. 2015) by mixing the WS mutant 1:1 with a reference strain labelled with green fluorescent protein and measuring the ratio of each strain before and after static growth for 24 h using flow cytometry (BD FACS Canto). We used a WspF ΔT226-G275 deletion mutant as the reference strain because WspF mutants are the most commonly found WS type when grown under selective conditions (McDonald, et al. 2009) and the in frame deletion of 50 amino acids most likely represents a complete loss-of-function mutation with minimal polar effects on the downstream wspR. Selection coefficients per generation were calculated as $s = \frac{\ln(R(t)/R(0))}{t}$, as previously described (Dykhuizen 1990) where $R$ is the ratio of alternative WS mutant to WspF ΔT226-G275 GFP and $t$ the number of generations. Viable counts on KB plates of initial and final populations were performed to calculate the number of generations. Stability of colony morphologies were confirmed and data from microcosms with >5% smooth colonies were excluded (two cases). Control competition experiments with isogenic WspF ΔT226-G275 reference strains with and without GFP were used to correct for the cost of the GFP marker. Control competitions were also used to determine the cost of the double deletions and the reporter construct relative to a wild type genetic background, for example an AwsX ΔY77-Q87 mutant in Δwsp/Δmws background
with pMSC was competed with a GFP labeled AwsX ΔY77-Q87 mutant in wild type background. Competitions were performed in independently inoculated quadruplicates for each strain with the number of replicates based on previous work (Lind, et al. 2015).

**Homology models**

Homology models of the structure of WspA, WspE, WspR, AwsR, AwsO and MwsR were made using Phyre2 in intensive mode (http://www.sbg.bio.ic.ac.uk/phyre2) (Kelley, et al. 2015).

**Probability estimation in the null Model IV**

The differential equation models describe the interactions between proteins in each of the three WS pathways. In order to solve the differential equations, two pieces of information are required: i) the initial concentrations of the molecular species and ii) the reaction rates. Although this information is unavailable a random-sampling approach was used to generate different random sets of initial concentrations and reaction rates. Each random set was used to establish a baseline of potential WS expression making it possible to evaluate whether a set of mutations results in a WS type. Effectively, this approach allows sampling of the probability distribution \( P(WS \mid m_i \in Wsp) \) used in our Bayesian model.

We randomly sample 1,000 different sets of reaction rates and initial concentrations from uniform priors: reaction rates were sampled randomly from a uniform distribution on log space (i.e. \( 10^{U[-2,2]} \)) and initial concentrations of reactants were sampled from a uniform distribution \( U[0,10] \). For each set, the appropriate differential
equation model was integrated and the steady state concentration of the compounds that correspond to a wrinkly spreader (RR in Aws, R* in Wsp and D* for Mws) computed. This served as a baseline for the non-WS phenotype that was used for comparison to determine whether combinations of mutations result in increased WS expression. After obtaining the baseline, we implemented particular combinations of enabling/disabling mutations \( (a_m) \). Ideally, a distribution linking enabling/disabling mutations to a fold change in reaction rates would be used, but this information is unavailable. In order to progress the effect sizes for enabling and disabling mutations were sampled from \( 10^{U[0,2]} \) and \( 10^{U[-2,0]} \), respectively, and then multiplied by the reaction rates. The differential equations were then solved for the same time that it took the baseline to reach steady state. The final concentration of R* (Figure 4A), RR (Figure 4B) and D* (Figure 4C) was then compared to the baseline and the number of times out of 1,000 that the WS-inducing compound increased served as an estimate of \( P(WS|m_i \in Wsp) \). The probability distribution stabilized by 500 random samples and additional sampling did not produce significant changes (data not shown).

The absence of empirical data on reaction rates, initial concentrations, and expected mutation effect size meant using a random sampling approach requiring estimates for parameter ranges. Parameter ranges were chosen to be broad enough to capture differences spanning several orders of magnitudes while allowing numerical computations for solving the differential equations. To assess the effect of these ranges on the results, the sampling procedure was repeated for WSP for three different parameter regimes i) an expanded range for initial concentrations [0-50], ii) an expanded range for reaction rates \( 10^{[3,3]} \), iii) a compressed range for mutational
effect size $10^{A^{-1.1}}$. This analysis shows that qualitative results are robust to these
changes (see Figure S1).

Supplementary information

Figure 6 – source data. Table of all WS mutations in Wsp, Aws and Mws.
Figure 9 – source data. Data from fitness assay.

Supplementary text: Analysis of mutations and molecular effects. Source code, Figure S1

Acknowledgements

This work was supported by Marsden Fund Council from New Zealand Government funding, administered by the Royal Society of New Zealand.

Competing interests

The authors declare no competing interests.

Author contributions

Peter A. Lind Conceptualization, Methodology, Investigation, Data Curation, Writing—original draft, Writing—review and editing, Visualization

Eric Libby Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing—original draft, Writing—review and editing, Visualization

Jenny Herzog Methodology, Investigation, Data Curation
Paul B. Rainey Conceptualization, Methodology, Validation, Writing—original draft, Writing—review and editing, Visualization, Supervision, Project administration, Funding acquisition

References

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

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

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

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

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

Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. 2013. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. BMC Genomics 14 Suppl 3:S6. http://dx.doi.org/10.1186/1471-2164-14-S3-S6
Capriotti E, Fariselli P, Casadio R. 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 33:W306-310. http://dx.doi.org/10.1093/nar/gki375

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

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

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

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

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

De N, Navarro MV, Raghavan RV, Sondermann H. 2009. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. J Mol Biol 393:619-633. http://dx.doi.org/10.1016/j.jmb.2009.08.030
de Visser JA, Krug J. 2014. Empirical fitness landscapes and the predictability of evolution. Nature reviews. Genetics 15:480-490. http://dx.doi.org/10.1038/nrg3744

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

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

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

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

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

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

Flowers JM, Hanzawa Y, Hall MC, Moore RC, Purugganan MD. 2009. Population genomics of the Arabidopsis thaliana flowering time gene network. Molecular biology and evolution 26:2475-2486. http://dx.doi.org/10.1093/molbev/msp161
Foster PL, Lee H, Popodi E, Townes JP, Tang H. 2015. Determinants of spontaneous mutation in the bacterium Escherichia coli as revealed by whole-genome sequencing. Proceedings of the National Academy of Sciences of the United States of America 112:E5990-5999. http://dx.doi.org/10.1073/pnas.1512136112

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

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

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

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

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

Gompel N, Prud'homme B. 2009. The causes of repeated genetic evolution. Dev Biol 332:36-47. http://dx.doi.org/10.1016/j.ydbio.2009.04.040
Gould SJ. 1989. Wonderful life: the Burgess Shale and the nature of history. New York: W.W. Norton.

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

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

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

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

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

Herron MD, Doebeli M. 2013. Parallel evolutionary dynamics of adaptive diversification in Escherichia coli. PLoS biology 11:e1001490. http://dx.doi.org/10.1371/journal.pbio.1001490
Hietpas RT, Jensen JD, Bolon DN. 2011. Experimental illumination of a fitness landscape. Proc Natl Acad Sci U S A 108:7896-7901.
http://dx.doi.org/10.1073/pnas.1016024108

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

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

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

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

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

King EO, Ward MK, Raney DE. 1954. Two simple media for the demonstration of pyocyanin and fluorescin. The Journal of laboratory and clinical medicine 44:301-307.
Kouyos RD, Leventhal GE, Hinkley T, Haddad M, Whitcomb JM, Petropoulos CJ, Bonhoeffer S. 2012. Exploring the complexity of the HIV-1 fitness landscape. PLoS genetics 8:e1002551. http://dx.doi.org/10.1371/journal.pgen.1002551

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

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

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

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

Laland KN, Uller T, Feldman MW, Sterelny K, Muller GB, Moczek A, Jablonka E, Odling-Smee J. 2015. The extended evolutionary synthesis: its structure, assumptions and predictions. Proceedings. Biological sciences / The Royal Society 282:20151019. http://dx.doi.org/10.1098/rspb.2015.1019
Lind PA, Andersson DI. 2008. Whole-genome mutational biases in bacteria. Proceedings of the National Academy of Sciences of the United States of America 105:17878-17883. http://dx.doi.org/10.1073/pnas.0804445105

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

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

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

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

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

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

Lundin E, Tang PC, Guy L, Nasvall J, Andersson DI. 2017. Experimental determination and prediction of the fitness effects of random point mutations in the
biosynthetic enzyme HisA. Molecular biology and evolution.

http://dx.doi.org/10.1093/molbev/msx325

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

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

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

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

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

Martin P, Sun L, Hood DW, Moxon ER. 2004. Involvement of genes of genome maintenance in the regulation of phase variation frequencies in Neisseria 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
Orgogozo V. 2015. Replaying the tape of life in the twenty-first century. Interface Focus 5:20150057. http://dx.doi.org/10.1098/rsfs.2015.0057

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

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

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

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

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

Pigliucci M. 2010. Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 365:557-566. http://dx.doi.org/10.1098/rstb.2009.0241
Pollock DD, Larkin JC. 2004. Estimating the degree of saturation in mutant screens. Genetics 168:489-502. http://dx.doi.org/10.1534/genetics.103.024430

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

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

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

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

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

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

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

Rosche WA, Foster PL. 2000. Determining mutation rates in bacterial populations.
Methods 20:4-17. http://dx.doi.org/10.1006/meth.1999.0901
Sanjuan R. 2010. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. Philos Trans R Soc Lond B Biol Sci 365:1975-1982. http://dx.doi.org/10.1098/rstb.2010.0063

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

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

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

Shewaramani S, Finn TJ, Leahy SC, Kassen R, Rainey PB, Moon CD. 2017. Anaerobically Grown Escherichia coli Has an Enhanced Mutation Rate and Distinct Mutational Spectra. PLoS genetics 13:e1006570. http://dx.doi.org/10.1371/journal.pgen.1006570
Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C. 2005. Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. Plant physiology 138:1163-1173. http://dx.doi.org/10.1104/pp.105.061309

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

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

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

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

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

Stern DL. 2013. The genetic causes of convergent evolution. Nature reviews.

Genetics 14:751-764. http://dx.doi.org/10.1038/nrg3483
Stern DL, Orgogozo V. 2009. Is genetic evolution predictable? Science 323:746-751. 
http://dx.doi.org/10.1126/science.1158997

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

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

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

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

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

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

Weinreich DM, Delaney NF, Depristo MA, Hartl DL. 2006. Darwinian evolution can 
follow only very few mutational paths to fitter proteins. Science 312:111-114. 
http://dx.doi.org/10.1126/science.1123539
Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE. 2006. Tests of parallel molecular evolution in a long-term experiment with Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America 103:9107-9112. http://dx.doi.org/10.1073/pnas.0602917103

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

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

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

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