Disentangling the effects of genetic architecture, mutational bias and selection on evolutionary forecasting

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

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

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

email: peter.lind@umu.se

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

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

Impact statement: Using a combination of genetics, experimental evolution and mathematical modelling this work 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 routes that evolution follows and the expected mutational targets. To test predictions, mutation rates and targets were determined for each pathway. Unanticipated mutational hotspots caused data to depart from predictions but the new data were readily incorporated into 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 maybe, 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 (Luksza and Lassig 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 (Pigliucci 2010; Stern 2013; Laland, et al. 2015). This is particularly apparent in 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).

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, but these factors alone leave unconsidered mechanistic details that arise from the genotype-to-phenotype map and from mutational biases. To what extent do these details matter?

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 (Alberch 1991; Gompel and Prud'homme 2009; Stern and Orgogozo 2009; Rainey, et

al. 2017). The function of gene products and their regulatory interactions thus provides information on likely mutational targets underpinning particular phenotypes. This is evident when considering 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 Stojiljkovic 2001; Martin, et al. 2004; Hammerschmidt, et al. 2014; Heilbron, et al. 2014).

Certain palindromic structures also lead to mutational bias (Viswanathan, et al. 2000; Lovett 2004) and promote amplification events including that increase mutational target size (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 also skew the distributions of mutational effects. Further 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

112 environmental factors that affect not only mutation rate but also the spectra of
113 mutational events (Krasovec, et al. 2017; Maharjan and Ferenci 2017; Shewaramani,
114 et al. 2017).
115
116 Beyond the genotype-to-phenotype map and mutational biases, predicting adaptive
117 evolution requires ability to know *a priori* the fitness effects of specific mutations. At
118 the present time there is much theoretical and empirical interest in the distribution of
119 fitness effects (DFE) (Eyre-Walker and Keightley 2007) — and particularly the DFE
120 of beneficial mutations (Orr 2005) — because of implications for predicting the rate
121 of adaption and likelihood of parallel evolution (de Visser and Krug 2014), but
122 knowledge of the shape of the distribution is insufficient to connect specific mutations
123 to their specific effects, or to their likelihood of occurrence. Such connections require
124 a means of knowing the connection between mutations and their environment-specific
125 fitness effects. This is tall order. A starting point is to understand the relationship
126 between all possible mutational routes to a particular phenotype and the set that are
127 realised by selection.
128
129 Here we take a bacterial system in which the genetic pathways underpinning
130 evolution of the adaptive "wrinkly spreader" (WS) type are known and use this to
131 explore the current limits on evolutionary forecasting. *Pseudomonas fluorescens*
132 SBW25 growing in static broth microcosms rapidly depletes available oxygen
133 establishing selective conditions that favour mutants able to colonise the air-liquid
134 interface. The most successful mutant-class encompasses the WS types (Ferguson, et
135 al. 2013; Lind, et al. 2017b). These types arise from mutational activation of
136 diguanylate cyclases (DGCs), cause over-production of the second messenger c-di-
137 GMP (Goymer, et al. 2006; McDonald, et al. 2009), over-production of an acetylated
138 cellulose polymer (Spiers, et al. 2002; Spiers, et al. 2003) and ultimately formation of
139 a self-supporting microbial mat (Figure 1A).

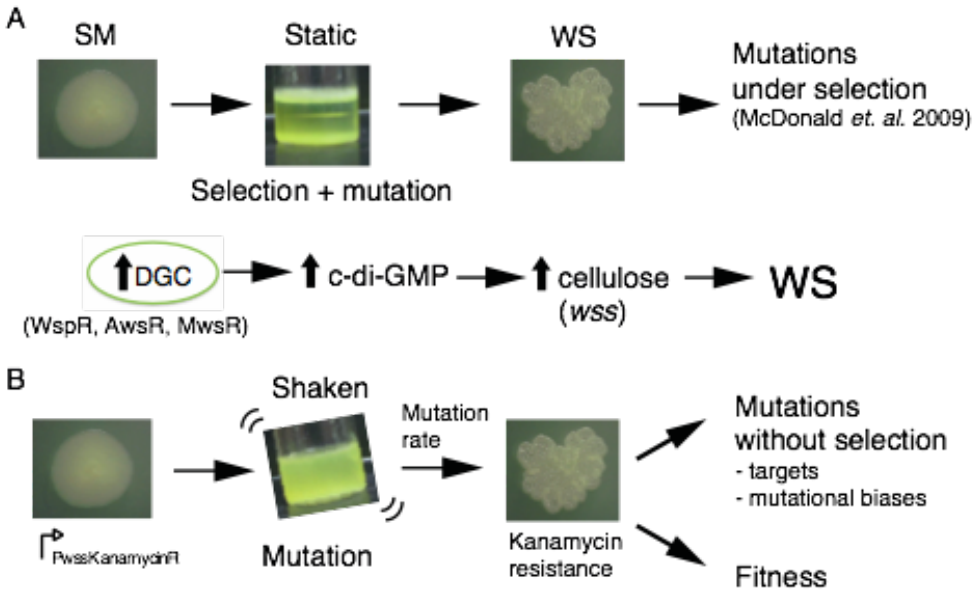


Figure 1. The *Pseudomonas fluorescens* SBW25 wrinkly spreader model. (**A**) Selection for access to oxygen allows wrinkly spreader (WS) mutants to invade the ancestral smooth (SM) population in static microcosms. WS mutants form a mat at the air-liquid interface through increased expression of the main structural component, cellulose, encoded by the *wss* operon. Expression of cellulose is controlled by the second messenger c-di-GMP, which is produced by diguanylate cyclases (DGCs). Mutations in the *wsp*, *aws* and *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 *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 and the influence of fitness, mutational target size and mutational biases on the outcome of evolution can be evaluated.

McDonald *et al.* (McDonald, 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 that evolution proceeds along multiple new pathways (Lind, et al. 2015). Preferential usage of Wsp, Aws and Mws pathways stems from the fact that they 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.

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 feasibility of developing bottom-up strategies for evolutionary forecasting. Our findings show that mechanistic-level predictions 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 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 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.

The mutation rate was highest for the Aws pathway (6.5×10^{-9}); approximately double that of Wsp (3.7×10^{-9}) and an order of magnitude higher than that of the Mws pathway (0.74×10^{-9}) (Figure 2). The rate at which WS mutants arose from the ancestral genotype in which the three pathways are intact (11.2×10^{-9}) was approximately the sum of the rates for the three pathways (11.0×10^{-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 compared to seven genes in the Wsp pathway).

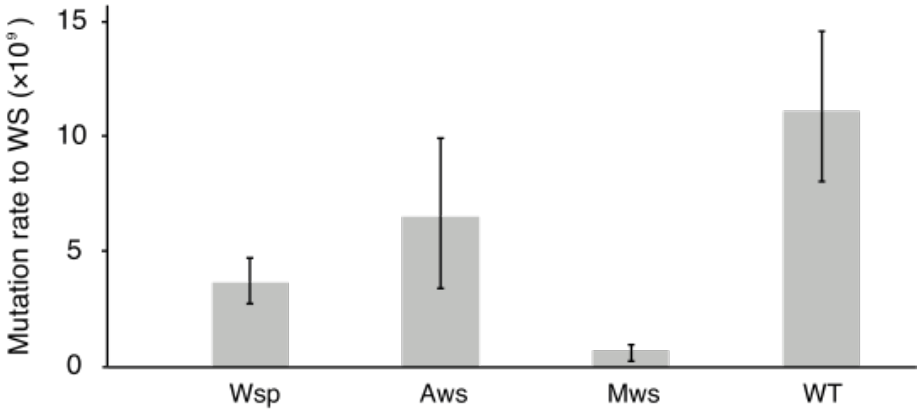


Figure 2. Mutation rates to WS. Fluctuation tests were used to estimate the mutation rate to WS for the three common mutational pathways to WS. Error bars represent mean and 95% confidence intervals.

Modelling 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 pathways. 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; Romling, et al. 2013; Micali and Endres 2016) comprised 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 WspF methylesterase WspF (PFLU1224) following transference of an active phosphoryl group. The activity of WspA is modulated by methylation: the constitutively active CheR-like methyltransferase WspC (PFLU1221) transfers methyl groups to conserved glutamine residues on WspA while when phosphorylated, WspF serves to remove these groups. 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, known as *morA* (PA4601) in *Pseudomonas aeruginosa*), 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).

If the specific effect of changing each nucleotide (and sets of nucleotides) was known then models for each pathway would not be required, but such knowledge does not exist. We therefore take a simplifying approach in which attention focuses on the interactions between components that correspond to reactions whose rate can either increase, decrease, or remain unaffected, depending on mutations in the component parts. Such mutations increase the reaction (enabling mutations), decrease the reaction (disabling mutations) or leave them unaffected. The components and interactions are shown in Figure 3. Figure 3 along with figure supplements 1 to 3 describe the molecular reactions and the associated differential equations governing the dynamics of each pathway. An advantage of this simplifying approach is that changes to a reaction may encompass mutations in more than a single component. For example, mutations in either WspE or WspR may increase reaction r_5 of the Wsp pathway (Figure 3A).

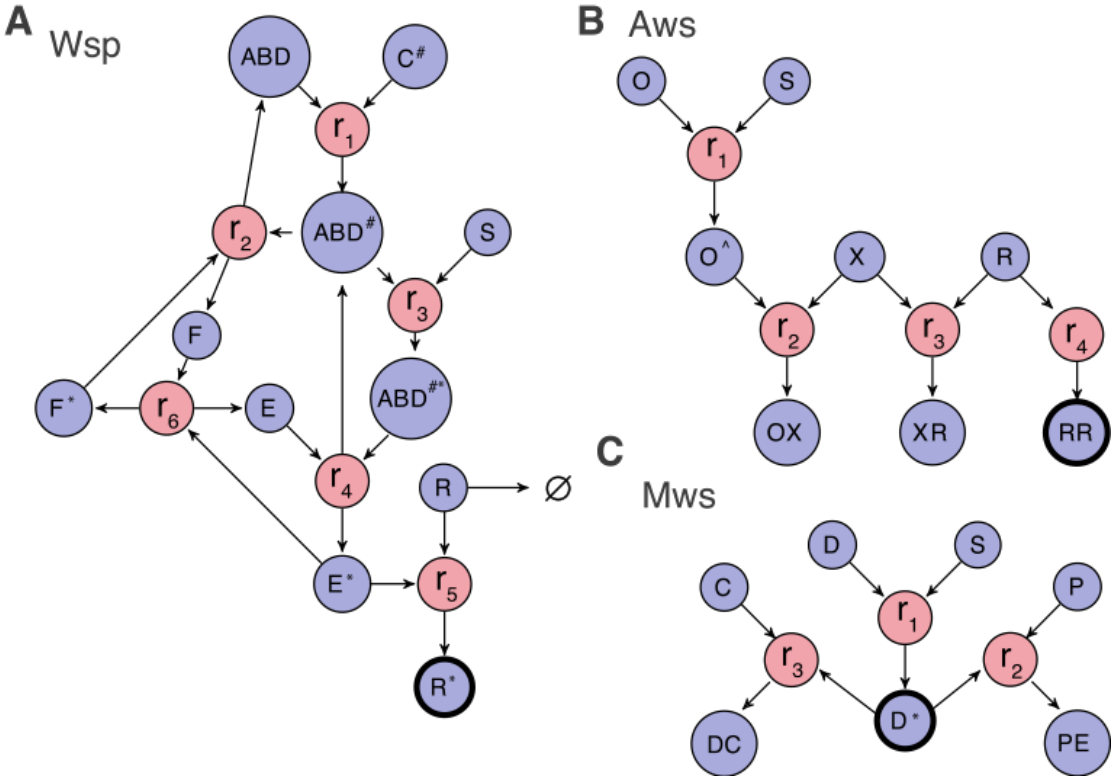


Figure 3. Modeling of WS. (A) Model of the Wsp pathway. R* represents the activated form of WspR and increase of R* leads to a WS phenotype **(B)** Model for the Aws pathway. RR represents the activated form of AwsR and increase of RR leads to a WS phenotype. (**C**) Model of the Mws Pathway. D* represents the activated form of the DGC domain. Our functional model places the PDE domain as a negative regulator of DGC activity. Details of the molecular networks are found in Figure 3 figure supplement 1 for Wsp, Figure 3 figure supplement 2 for Aws and Figure 3 figure supplement 3 for Mws. In the simple null models all genetic components are the same size, but information of mutational target size can readily be accommodated in the model by changing the individual probabilities of disabling and enabling mutations. The benefit of doing so might however be small if not combined with detailed data on mutation rates (see Discussion).

Equipped with the set of mathematical descriptions it is possible to consider combinations of enabling, disabling, and no effect changes to reaction rates and determine the likelihood that a WS type is generated. For the Wsp system this amounts to 3^6 or 729 combinations. 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, or r_6 (Figure 3A)).

Predicting the pathways that evolution follows and genetic targets

To determine whether mutations producing WS occur more often in Wsp compared to Aws or Mws pathways, we adopt a Bayesian approach in which the probability that a particular pathway is 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).

$$P(WS \cap m \cap Wsp) = \sum_i P(WS|m_i \in Wsp)P(m_i \in Wsp) \qquad (1)$$

To estimate P($m_i \in$ Wsp) we assume fixed probabilities of enabling and disabling mutations 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), requires knowing both how gene products interact and how these interactions result in a phenotype. This information is estimated based on the pathway dynamics represented in Figure 3 and Figure 3 – figure supplement 1 by repeated sampling from the space of all possible reaction rates, initial concentrations, and magnitude of effects (see Materials and Methods).

The results of simulations are shown in Figure 4. Figure 4A shows that the Wsp pathway is predicted to be the target of mutation 1.3 - 2.1 times more often than the Aws pathway while Figure 4B shows that the Mws pathway is predicted to be the target of mutation 0.7 - 1.0 times less often that 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 mutation rate data showing WS types to be twice as likely to arise from mutation in Aws, versus Wsp. The causes of this discrepancy are described in the following section.

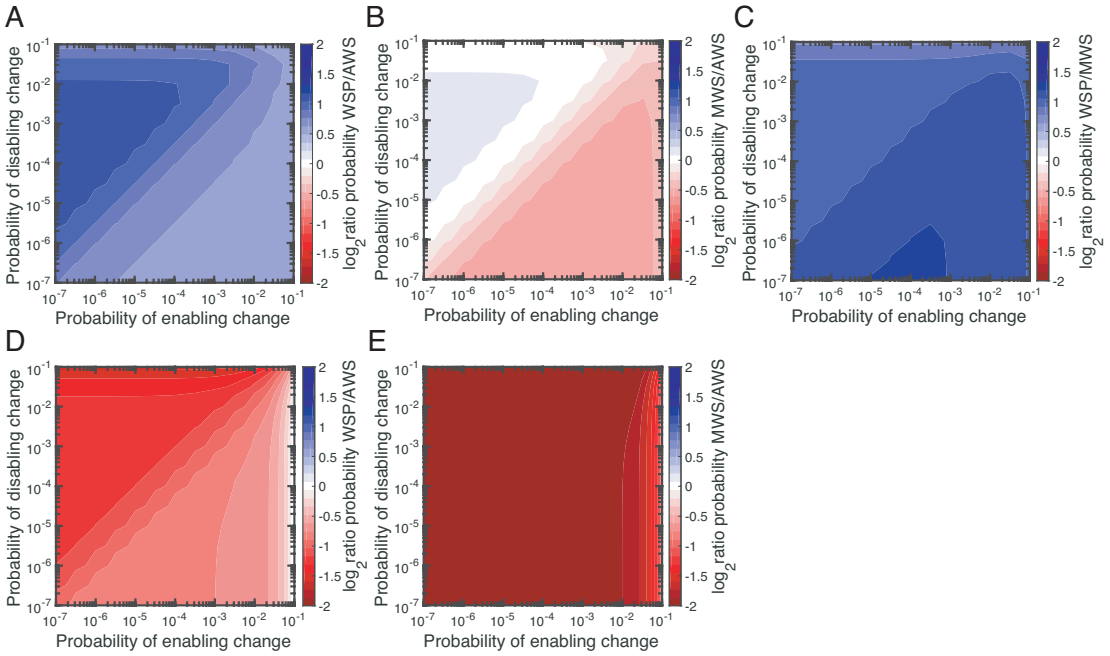


Figure 4. Modelling of the probability of using the Wsp, Aws and Mws pathways

(A) Probability of Wsp relative to Aws. Equal mutation rates for all components.

(**B**) **Probability of Mws relative to Aws.** Equal mutation rates for all components.

(C) Probability of Mws relative to Wsp. Equal mutation rates for all components.(**D**) **Probability of Wsp relative to Aws with the mutational hotspot in AwsX included.** Based on the mutational data a hot spot for Aws is included in the models that increase the mutation rate by a factor five for both enabling and disabling mutations. (**E**) **Probability of Mws relative to Aws with the mutational hotspot in AwsX included.** The relative contributions of individual reactions rates are available in Figure 4 - figure supplement 1 for Wsp and Figure 4 - figure supplement 1 for Aws.

In addition to predicting the preferred mutational pathways to produce a WS, the Bayesian approach also predicts genes likely to be affected by mutation (Figure 4 figure supplement 1 and 2). Predictions as to specific genetic targets come from appraisal of the relative importance of each reaction in terms of generating a wrinkly spreader (Figure 4 – supplement 1 and 2). While it is recognised that a majority of WS mutations arise from mutations in negative regulators of DGC activity, 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 4 – figure supplement 1), there are two reactions (r_2 and r_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 phosphoric 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 4 – 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 4 – 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 4 – figure supplement 1). Enabling mutations based on the model 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 2A). 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 3B, Figure 4 – 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 5A, Figure 5 – 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).

The distribution of mutations for each of the three pathways is indicative of bias. As shown in Figure 5B, 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 can 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 5B).

The mathematical models presented above assumed no mutational bias thus the lack of fit between mutation rate data and predictions from the models. Nonetheless, changing specific reaction rates within the models readily incorporates such knowledge. For example, the mutational hotspot in *awsX* affects reactions r_2 and r_3 in the Aws differential equation system (Figure 3B, Figure 3 – figure supplement 2). The effect of a five-fold change in the probability of enabling/disabling change in these reactions leads to the prediction that the Aws pathway is more likely to generate WS types than Wsp for most probability values (see Figure 4D). The only area of parameter space in which evolution is more likely to utilise the Wsp pathway is for rare mutations that have a high probability of enabling change ($>> 10^{-2}$). One interesting consequence is that it changes the phase-space over which evolution of WS via mutations in the Wsp pathway is more likely with respect to the Aws pathway. In Figure 4A, evolution is most likely to proceed via the Wsp pathway when the probability of disabling change is greater than the probability of enabling change. In contrast, when the likelihood of producing WS types is affected by the mutational hotspot in *awsX*, then evolution will proceed via Wsp only when the probability of enabling change is greater than the probability of disabling change (Figure 4D and Figure 4 figure supplement 2).

Analysis of mutants reveals mutational targets and effects

Wsp pathway: Mutations were identified in five genes of the seven-gene pathway all of which were predicted by the null model (Figure 4 – figure supplement 1). The most commonly mutated gene was *wspA* (PFLU1219), with ten of 15 mutations (Figure 5) being amino acid substitutions (six unique) clustered in the region 352-420 at the stalk of the signalling domain. This region has been implicated in trimer-of-dimer formation for the WspA homologue in *Pseudomonas aeruginosa* (O'Connor, et al. 2012) which is critical for self-assembly and localization of Wsp clusters in the membrane. It is possible that these mutations stabilize trimer of dimer formation,

change the subcellular location of the Wsp complex, or affect interaction with WspD (putative interface 383-420 in WspA) (Griswold, et al. 2002) and thus affecting relay of signal to WspE. These effects we interpreted as enabling mutations increasing r_3 in Figure 2A. The four additional mutations were in frame deletions in a separate region of the transducer domain (ΔT293 - E299, ΔA281-A308). Although it is possible that these mutations could also affect trimer-of-dimer formation, there are predicted methylation sites in the region (Rice and Dahlquist 1991) that regulate the activity of the protein via methyltransferase WspC and methylesterase WspF. Given that disabling mutations are more common than enabling mutations it is likely that these mutations decrease r_2 in Figure 3A by disrupting the interaction with WspF. We also identified a single mutation that fused the open reading frame of WspC, the methyltransferase that positively regulates WspA activity, to WspD, resulting in a chimeric protein (Figure 5, Figure 5 – source data). This mutation is likely to be a rare enabling mutation that increases the activity of WspC (increasing r_1 in Figure 3A) by physically tethering it to the WspABD complex thus allowing it to more effectively counteract the negative regulator WspF. Alternatively, the tethering may physically block the interaction with WspF (decrease of r_2 in Figure 3A).

The second most commonly mutated gene in the *wsp* operon was *wspE* (PFLU1223) (Figure 5). Four amino acids were repeatedly mutated in the response regulatory domain of WspE and all cluster closely in a structural homology model made with Phyre2 (Kelley, et al. 2015). All mutated residues surround the active site of the phosphorylated D682 and it is likely that they disrupt feedback regulation by decreasing phosphorylation of the negative regulator WspF (decreasing r_6) rather than increasing activation of WspR (r_5 in Figure 3A).

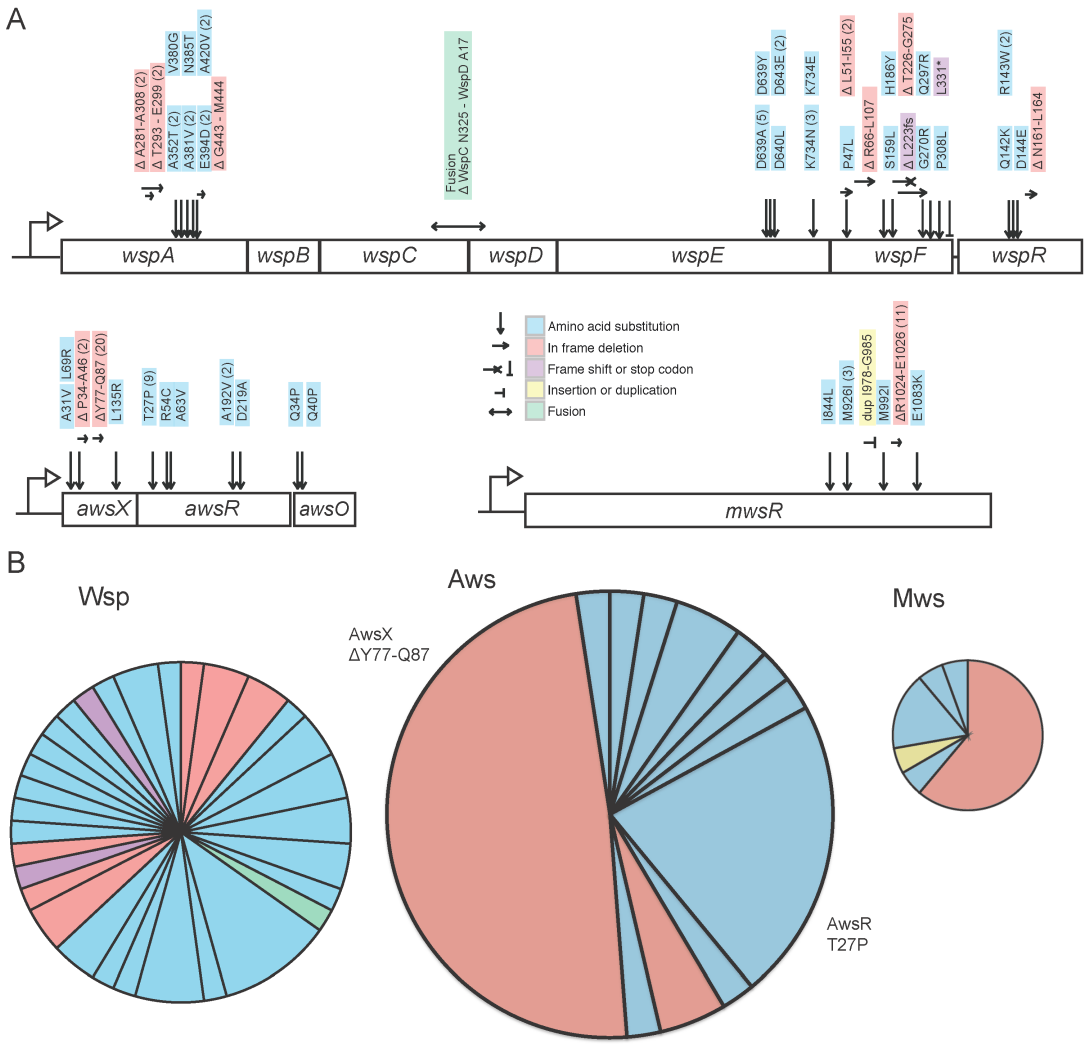


Figure 5. Mutational targets. (**A**) 105 independent mutations in the *wsp* (n=46), *aws* (n=41) and *mws* (n=18) operons were identified. Numbers of independent mutants are shown in brackets. Full details on the mutations are available in Figure 5 – source data 1. (**B**) Diversity of mutations with 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 increased mutation rate can be incorporated into the null model by increasing the probability of change for r_2 and r_3 in the Aws model (Figure 3B).

Twelve mutations were detected in *wspF* (PFLU1224). These are distributed throughout the gene and include amino acid substitutions, in-frame deletions as well as a frame-shift and a stop codon (Figure 5). The pattern of mutations is consistent with both the role of WspF as a negative regulator of WspA activity and the well-

characterised effect of loss-of-function mutations in this gene (Bantinaki, et al. 2007; McDonald, et al. 2009). The mutations are interpreted as decreasing r_2 in Figure 3A. Five mutations were found in WspR (PFLU1225), the DGC output response regulator that produces c-di-GMP and activates expression of cellulose (Figure 5). All mutations were located in the linker region between the response regulator and DGC domains. Mutations in this region are known to generate constitutively active *wspR* alleles by relieving the requirement for phosphorylation (Goymer, et al. 2006). They may additionally affect subcellular clustering of WspR (Huangyutitham, et al. 2013) or shift the equilibrium between the dimeric form of WspR, with low basal activity, towards a tetrameric activated form (De, et al. 2009). In our model these increase reaction r_5.

Aws pathway: Mutations were identified in all three genes of the Aws pathway – all of which were predicted by the null model. In the Aws pathway, mutations were most commonly found in *awsX* (25 out of 41 mutations (Figure 5)). The above-mentioned mutational hotspot produced in-frame deletions likely mediated by 6 bp direct repeats (Figure 5 – source data 1). The deletions are consistent with a loss of function and a decrease in r_3 (Figure 3B) that would leave the partially overlapping open reading frame of the downstream gene (*awsR*) unaffected.

The DGC AwsR, was mutated in 14 cases with an apparent mutational hot spot at T27P (9 mutants) in a predicted transmembrane helix (amino acids 19-41). The remaining mutations were amino acid substitutions in the HAMP linker and in the PAS-like periplasmic domain between the two transmembrane helices. These amino acid substitutions are removed from the output DGC domain (Figure 5) and their effects are difficult to interpret, but they could cause changes in dimerization (Malone, et al. 2012) or the packing of HAMP domains, which could, in turn, alter transmission of conformational changes in the periplasmic PAS-like domain to the DGC domain causing constitutive activation (Parkinson 2010). Such effects would increase r_4 in Figure 3B. Mutations in the N-terminal part of the protein are easier to interpret based on the existing functional model (Malone, et al. 2012) and most likely disrupt interactions with the periplasmic negative regulator AwsX resulting in a decrease in r_3 in Figure 3B.

Two mutations were found in the outer membrane lipoprotein protein AwsO between the signal peptide and the OmpA domain (Figure 5). Both mutations were glutamine to proline substitutions (Q34P, Q40P), which together with a previously reported G42V mutation (McDonald, et al. 2009) suggest that multiple changes in this small region can cause a WS phenotype. This is also supported by data from *Pseudomonas aeruginosa* in which mutations in nine different positions in this region lead to a small colony variant phenotype similar to WS (Malone, et al. 2012). A functional model based on the YfiBNR in *P. aeruginosa* (Malone, et al. 2012; Xu, et al. 2016), suggest that AwsO sequesters AwsX at the outer membrane and that mutations in the N-terminal part of the protein lead to constitutive activation and increased binding of AwsX. This would correspond to an increase in r_2 in Figure 3B, which would relieve negative regulation of AwsR.

Mws pathway: The MwsR pathway (comprising just a single gene) harbours mutations in both DGC and phosphodiesterase (PDE) domains . Only mutations in the C-terminal phosphodiesterase (PDE) domain were predicted (Figure 3C). Eleven of 18 mutations were identical in-frame deletions (ΔR1024-E1026) in the PDE domain, mediated by 8 bp direct repeats (Figure 5, Figure 5 – source data 1). It has been shown previously that deletion of the entire PDE domain generates the WS phenotype (McDonald, et al. 2009), suggesting a negative regulatory role that causes a decrease of r_2 in the model in Figure 3C. One additional mutation was found in the PDE domain (E1083K) located close to R1024 in a structural homology model made with Phyre2 (Kelley, et al. 2015), but distant to the active site residues (E1059-L1061). Previously reported mutations (A1018T, ins1089DV) (McDonald, et al. 2009) are also removed from the active site and cluster in the same region in a structural homology model. This suggests that loss of phosphodiesterase activity may not be the mechanism leading to the WS phenotype. This is also supported by the high solvent accessibility of the mutated residues, which indicates that major stability-disrupting mutations are unlikely and changes in interactions between domains or dimerization are more probable. Thus, it is likely that the WS phenotype resulting from a deletion in the PDE domain is caused by disruption of domain interactions or dimerization rather than loss of phosphodiesterase activity.

The remaining mutations within *mwsR* are amino acid substitutions in the GGDEF domain, close to the DGC active site (927-931) with the exception of a duplication of I978-G985. While it is possible that these mutations directly increase the catalytic activity of the DGC, increasing r_1 in Figure 3C, such enabling mutations are considered to be rare. An alternative hypothesis is that these mutations either interfere with c-di-GMP feedback regulation or produce larger conformational changes that change inter-domain or inter-dimer interactions, similar to the mutations in the PDE domain. Based on these data we reject the current model of Mws function, which predicted mutations decreasing r_2 (Figure 3C) through mutations inactivating the PDE domain. We instead suggest that the mutations are likely to disrupt the conformational dynamics between the domains and could be seen either as activating mutations causing constitutive activation or disabling mutations with much reduced mutational target size that must specifically disrupt the interaction surface between the domains. In both cases the previous model lead to an overestimation of the rate to WS for the Mws pathway.

Fitness of WS types

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 6). This type of fitness data should be interpreted with caution because the fitness of WS mutants have been shown to be frequency dependent and some WS mutants are superior in early phase attachment as opposed to growth at the air-liquid interface (Lind, et al. 2015). Nevertheless, these 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).

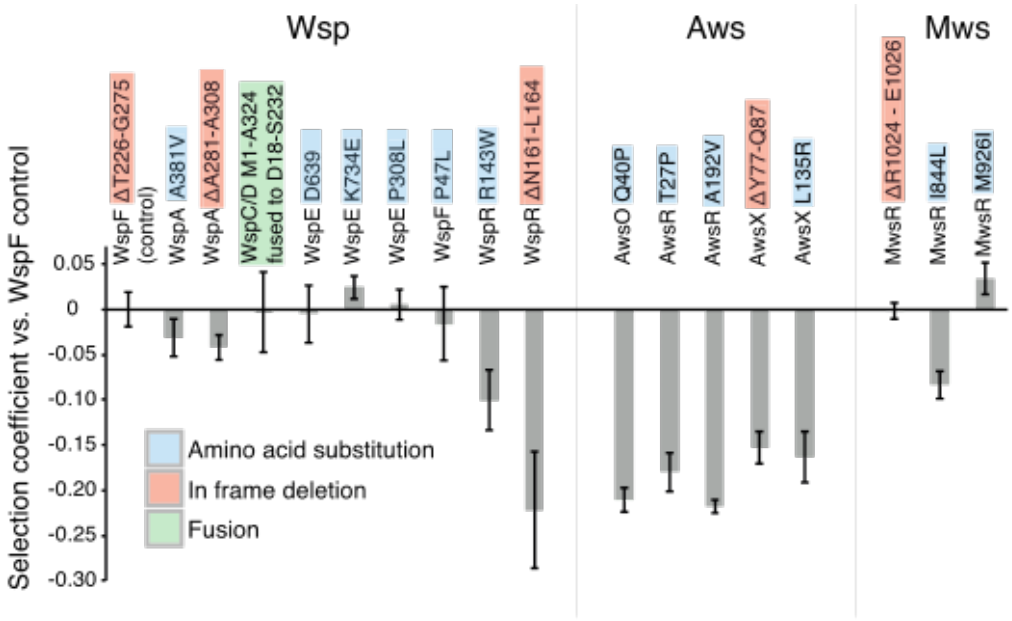


Figure 6. Fitness of different WS mutants. Competitive fitness against a WspF ΔT226-G275 reference strain was measured for representative mutations in the Wsp, Aws, Mws pathways. Pairwise competitions were performed in quadruplicates and error bars represent +/- one standard deviation.

The three *wspF* mutants, the *wspC-wspD* fusion, and the *wspE* mutants have similar fitness. In contrast, both *wspA* mutants are slightly less fit and both *wspR* mutants are severely impaired (Figure 6). 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 low fitness compared to the *wspF* reference strain (Figure 6), which explains 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.

Differences of mutational spectra with and without selection

A final question concerns the outcome of the original experimental evolution under selection (McDonald, et al. 2009) and whether it can be explained by our detailed measurement of mutation rates, mutational targets and fitness assays. As indicated above, there exist major differences in the spectrum of mutations isolated with and without selection (Figure 7). The most obvious difference is in the 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. The explanation lies in the lower fitness of Aws mutants (Figure 6). Similarly, fitness effects also explains differences in the spectrum of *wsp* mutations, with no *wspA* mutations being found under selection despite being the most commonly mutated gene without selection (15/46). The previous failure to detect *wspR* mutants in a screen of 53 WS mutants (Goymer, et al. 2006) is similarly explained by low fitness of WS types arising via mutations in this gene.

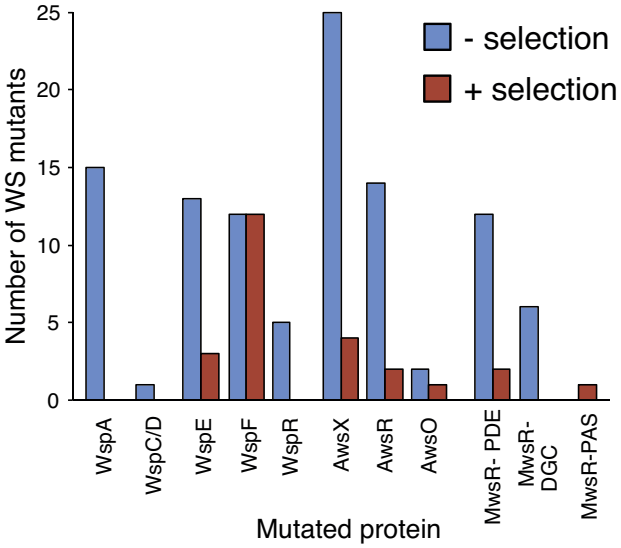


Figure 7. WS mutations isolated with and without selection. Fitness effects bias the mutational spectrum observed under selection resulting in underrepresentation of WspA and WspR compared to WspF and WspE. Similar fitness effects of different Aws mutants lead to similar patterns regardless of selective conditions. 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.

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 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 Doebeli

614 2013), but whether these cases are representative for evolutionary processes in
615 general remains to be determined. A related question, with greater potential for
616 practical applications, is whether it is possible to forecast short-term evolutionary
617 events and if so, then the challenge is to stipulate the data necessary to make
618 successful predictions.
619
620 Our uniquely detailed knowledge of the WS experimental evolution system has
621 provided a rare opportunity to disentangle the contributions of selection, mutational
622 biases and genetic architecture to evolutionary outcomes, and consequently explore
623 the limits of evolutionary forecasting. A thorough understanding of the function of the
624 molecular species and their interactions allowed development of a null model by
625 defining the genotype-to-phenotype map, which successfully predicted mutational
626 targets and the relative likelihood that evolution followed each of the three principle
627 pathways. Importantly, genetic architecture is likely to be transferable between
628 different species, which stands to allow the formulation of general principles and
629 evolutionary rules (Lind, et al. 2015). Despite the simplicity of the mathematical null
630 models, which contain only general information about functional interactions, we
631 successfully predicted mutational targets including previously unknown mutations in
632 *wspA*. Without information about fitness and mutational biases, however, only order
633 of magnitude predictions of mutant frequencies can be made. Thus, it is possible to
634 predict that Wsp, which is subject to negative regulation will be more common than a
635 DGC that requires enabling mutations (Lind, et al. 2015), but not which of two
636 pathways (Wsp and Aws) with differently wired negative regulation is likely to be
637 dominant after selection. Insights from the null model combined with data on
638 mutational targets also allowed us to reject our functional model of Mws.
639
640 Direct measurement of the fitness effects of large number of mutations is difficult and
641 time-consuming and typically only possible for microbial species. Therefore future
642 success in predicting fitness effects of mutations rests on the ability to infer them
643 from other parameters, such as estimated effects of mutations on thermodynamic
644 stability or molecular networks, or from incorporation of information concerning
645 evolutionary conservation of amino acid residues. Alternatively, it might be possible
646 to extrapolate findings from a small number of mutations that are either directly
647 constructed and assayed in the laboratory or through fitness estimates of

polymorphisms data for 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).

Interestingly WS mutations in the same gene often have similar fitness effects (Figure 6). Obviously no general conclusions can be drawn from these few cases, but if mutations with similar functional effects, for example disruption of a particular interaction, can be assumed to be equally fit, this would greatly reduce the number of specific mutants that need to be experimentally assayed for each gene. Several studies suggest that fitness distributions are often bimodal, with a significant proportion being complete loss-of-function mutations, which could explain the similar fitness effects of mutations in the same genes if they result in inactivation of a particular biochemical reaction of interaction (Sanjuan 2010; Jacquier, et al. 2013; Sarkisyan, et al. 2016; Lind, et al. 2017a). The extent to which fitness effects are transferable between strains with different genetic backgrounds or closely related species remains to be more fully investigated (Ungerer, et al. 2003; Pearson, et al. 2012; Wang, et al. 2014).

Estimates of genomic mutation rates are remarkably consistent across species (Drake 1991), and mutational biases as evident in the types of base substitutions, are well-characterized for a large number of bacterial species (Sung, et al. 2012; Wei, et al. 2014; Farlow, et al. 2015; Foster, et al. 2015; Long, et al. 2015). It is also known that molecular processes, such as transcription and replication, can introduce mutational biases (Beletskii and Bhagwat 1996; Hudson, et al. 2003; Lind and Andersson 2008; Reijns, et al. 2015; Zhao, et al. 2015) and mutational hotspots caused by homopolymeric tracts and direct repeats can greatly increase local mutation rate (Streisinger and Owen 1985; Seier, et al. 2011). However, the distribution of mutation rates across a gene or operon remains largely unknown. Absence of this knowledge currently hinders efforts to forecast adaptive evolution.

There are several cases of probable mutational hotspots in the spectrum of WS mutants found in this study before the influence of selection. One specific deletion (ΔY77-Q87) in *awsX* accounts 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 5) contribute equally to the evolution of WS types. Together, these findings draw attention to the limited value of including mutational target size alone as a parameter in predictive models.

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.

Evolutionary forecasting is likely to be most successful for biological systems where there are experimental data on a large number of independent evolutionary events, such as influenza, HIV and cancer (Kouyos, et al. 2012; Fraser, et al. 2014; Lawrence, et al. 2014; Luksza and Lassig 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 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

716 coarse and fine-grained approaches. Our demonstration that simple null models of
717 functional networks can produce highly relevant quantitative predictions is an
718 important step forward allowing predictions to be directly tested in other experimental
719 systems.
720

721 # Materials and methods

722

723 ## Strains and media

724 The strains used in the study are all *Pseudomonas fluorescens* SBW25 (Silby, et al.
725 2009) or derivatives thereof. The reporter construct (pMSC), used for isolation of WS
726 mutants before selection, fused the P_{wss} promoter to a kanamycin resistance marker
727 (*nptII*) (Fukami, et al. 2007; McDonald, et al. 2011). *P. fluorescens* strains with
728 deletions of the *wsp* (PFLU1219-1225), *aws* (PFLU5209-5211) and *mws* (PFLU5329)
729 operons were previously constructed as described by McDonald *et al.* (McDonald, et
730 al. 2011). All experiments used King's medium B (KB) (King, et al. 1954), solidified
731 with 1.5% agar and incubation was at 28°C. All strains were stored in glycerol saline
732 at -80°C.
733

734 ## Fluctuation tests and isolation of WS mutants before selection

735 Strains with the pMSC reporter construct and either wild type genetic background or
736 double deletions of *aws*/*mws*, *wsp*/*mws* or *wsp*/*aws* were used to estimate mutation
737 rates to WS before selection. Overnight cultures were diluted to approximately 10^3
738 cfu/ml and 60 independent 110 ul cultures were grown for 16-19 h (OD600= 0.9-1.0)
739 with shaking (200 rpm) in 96-well plates before plating on KB plates with 30 mg/l
740 kanamycin. Viable counts were estimated by serial dilution and plating on KB agar.
741 One randomly chosen colony per independent culture with WS colony morphology
742 was restreaked once on KB agar. The assay was repeated at least four times for the
743 double deletion mutants and twice for the wild type strain. Mutations rates were
744 estimated using the Ma-Sandri-Sarkar Maximum Likelihood Estimator (Hall, et al.
745 2009) available at www.keshavsingh.org/protocols/FALCOR.html.
746

747 ## 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 $s = [\ln(R(t)/R(0))]/[t]$, 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 determined using viable counts. 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 independently inoculated quadruplicates for each strain.

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 mathematical models

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 $|m_i \in$ 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_i). 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 3A), RR (Figure 3B) and D* (Figure 3C) 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 $10^{\wedge[-1,1]}$. This analysis shows that qualitative results are robust to these changes (see Figure S1).

Source code for the mathematical modelling is deposited as supplemental material.

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

842 # References

843

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

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

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

853 Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF.
854 2009. Genome evolution and adaptation in a long-term experiment with
855 Escherichia coli. Nature 461:1243-1247. http://dx.doi.org/10.1038/nature08480

856 Beletskii A, Bhagwat AS. 1996. Transcription-induced mutations: increase in C to T
857 mutations in the nontranscribed strand during transcription in Escherichia coli.
858 Proceedings of the National Academy of Sciences of the United States of America
859 93:13919-13924.

860 Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Jr., Dickson M, Grimwood
861 J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel
862 evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science
863 307:1928-1933. 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

Drake JW. 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proceedings of the National Academy of Sciences of the United States of America 88:7160-7164.

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

884 Farlow A, Long H, Arnoux S, Sung W, Doak TG, Nordborg M, Lynch M. 2015. The
885 Spontaneous Mutation Rate in the Fission Yeast Schizosaccharomyces pombe.
886 Genetics 201:737-744. http://dx.doi.org/10.1534/genetics.115.177329

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

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

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

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

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

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

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

913 Gompel N, Prud'homme B. 2009. The causes of repeated genetic evolution. Dev Biol
914 332:36-47. http://dx.doi.org/10.1016/j.ydbio.2009.04.040

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

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

921 Griswold IJ, Zhou HJ, Matison M, Swanson RV, McIntosh LP, Simon MI, Dahlquist
922 FW. 2002. The solution structure and interactions of CheW from Thermotoga
923 maritima. Nature Structural Biology 9:121-125. http://dx.doi.org/DOI
924 10.1038/nsb753

925 Hall BM, Ma CX, Liang P, Singh KK. 2009. Fluctuation analysis CalculatOR: a web
926 tool for the determination of mutation rate using Luria-Delbruck fluctuation

927 analysis. Bioinformatics 25:1564-1565.
928 http://dx.doi.org/10.1093/bioinformatics/btp253

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

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

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

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

941 Hudson RE, Bergthorsson U, Ochman H. 2003. Transcription increases multiple
942 spontaneous point mutations in Salmonella enterica. Nucleic Acids Res 31:4517-
943 4522.

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

945 Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, Bercot B, Petit
946 E, Poulain J, Barnaud G, et al. 2013. Capturing the mutational landscape of the
947 beta-lactamase TEM-1. Proc Natl Acad Sci U S A 110:13067-13072.
948 http://dx.doi.org/10.1073/pnas.1215206110

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

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

956 King EO, Ward MK, Raney DE. 1954. Two simple media for the demonstration of
957 pyocyanin and fluorescin. The Journal of laboratory and clinical medicine 44:301-
958 307.

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

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

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

970 Kugelberg E, Kofoid E, Andersson DI, Lu Y, Mellor J, Roth FP, Roth JR. 2010. The
971 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

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

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

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

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

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

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

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

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

1000 Lind PA, Farr AD, Rainey PB. 2015. Experimental evolution reveals hidden diversity
1001 in evolutionary pathways. Elife 4. http://dx.doi.org/10.7554/eLife.07074

1002 Long H, Kucukyildirim S, Sung W, Williams E, Lee H, Ackerman M, Doak TG,
1003 Tang H, Lynch M. 2015. Background Mutational Features of the Radiation-
1004 Resistant Bacterium Deinococcus radiodurans. Molecular biology and evolution
1005 32:2383-2392. http://dx.doi.org/10.1093/molbev/msv119

1006 Lovett ST. 2004. Encoded errors: mutations and rearrangements mediated by
1007 misalignment at repetitive DNA sequences. Mol Microbiol 52:1243-1253.
1008 http://dx.doi.org/10.1111/j.1365-2958.2004.04076.x

1009 Luksza M, Lassig M. 2014. A predictive fitness model for influenza. Nature 507:57-
1010 61. http://dx.doi.org/10.1038/nature13087

1011 Luria SE, Delbruck M. 1943. Mutations of Bacteria from Virus Sensitivity to Virus
1012 Resistance. Genetics 28:491-511.

1013 Lynch M. 2010. Rate, molecular spectrum, and consequences of human mutation.
1014 Proceedings of the National Academy of Sciences of the United States of America
1015 107:961-968. http://dx.doi.org/10.1073/pnas.0912629107

1016 Maharjan RP, Ferenci T. 2017. A shifting mutational landscape in 6 nutritional states:
1017 Stress-induced mutagenesis as a series of distinct stress input-mutation output
1018 relationships. PLoS biology 15:e2001477.
1019 http://dx.doi.org/10.1371/journal.pbio.2001477

1020 Malone JG, Jaeger T, Manfredi P, Dotsch A, Blanka A, Bos R, Cornelis GR, Haussler
1021 S, Jenal U. 2012. The YfiBNR signal transduction mechanism reveals novel targets
1022 for the evolution of persistent Pseudomonas aeruginosa in cystic fibrosis airways.
1023 PLoS Pathog 8:e1002760. http://dx.doi.org/10.1371/journal.ppat.1002760

1024 Malone JG, Jaeger T, Spangler C, Ritz D, Spang A, Arrieumerlou C, Kaever V,
1025 Landmann R, Jenal U. 2010. YfiBNR mediates cyclic di-GMP dependent small
1026 colony variant formation and persistence in Pseudomonas aeruginosa. PLoS
1027 Pathog 6:e1000804. http://dx.doi.org/10.1371/journal.ppat.1000804

1028 Martin P, Sun L, Hood DW, Moxon ER. 2004. Involvement of genes of genome
1029 maintenance in the regulation of phase variation frequencies in Neisseria
1030 meningitidis. Microbiology 150:3001-3012.
1031 http://dx.doi.org/10.1099/mic.0.27182-0

1032 McDonald MJ, Cooper TF, Beaumont HJE, Rainey PB. 2011. The distribution of
1033 fitness effects of new beneficial mutations in Pseudomonas fluorescens. Biology
1034 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

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

1057 Pearson VM, Miller CR, Rokyta DR. 2012. The consistency of beneficial fitness
1058 effects of mutations across diverse genetic backgrounds. Plos One 7:e43864.
1059 http://dx.doi.org/10.1371/journal.pone.0043864

1060 Phippen CW, Mikolajek H, Schlaefli HG, Keevil CW, Webb JS, Tews I. 2014.
1061 Formation and dimerization of the phosphodiesterase active site of the
1062 Pseudomonas aeruginosa MorA, a bi-functional c-di-GMP regulator. FEBS Lett
1063 588:4631-4636. http://dx.doi.org/10.1016/j.febslet.2014.11.002

1064 Pigliucci M. 2010. Genotype-phenotype mapping and the end of the 'genes as
1065 blueprint' metaphor. Philosophical transactions of the Royal Society of London.
1066 Series B, Biological sciences 365:557-566.
1067 http://dx.doi.org/10.1098/rstb.2009.0241

1068 Pollock DD, Larkin JC. 2004. Estimating the degree of saturation in mutant screens.
1069 Genetics 168:489-502. http://dx.doi.org/10.1534/genetics.103.024430

1070 Rainey PB, Remigi P, Farr AD, Lind PA. 2017. Darwin was right: where now for
1071 experimental evolution? Curr Opin Genet Dev 47:102-109.
1072 http://dx.doi.org/10.1016/j.gde.2017.09.003

1073 Reams AB, Roth JR. 2015. Mechanisms of gene duplication and amplification. Cold
1074 Spring Harb Perspect Biol 7:a016592.
1075 http://dx.doi.org/10.1101/cshperspect.a016592

1076 Reijns MA, Kemp H, Ding J, de Proce SM, Jackson AP, Taylor MS. 2015. Lagging-
1077 strand replication shapes the mutational landscape of the genome. Nature 518:502-
1078 506. http://dx.doi.org/10.1038/nature14183

1079 Rice MS, Dahlquist FW. 1991. Sites of deamidation and methylation in Tsr, a
1080 bacterial chemotaxis sensory transducer. J Biol Chem 266:9746-9753.

1081 Richardson AR, Stojiljkovic I. 2001. Mismatch repair and the regulation of phase
1082 variation in Neisseria meningitidis. Mol Microbiol 40:645-655.

1083 Roth JR, Benson N, Galitsky T, Haack K, Lawrence JG, Miesel L. 1996.
1084 Rearrangements of the Bacterial Chromosome: Formation and
1085 Applications. In: Neidhardt FC, editor. Escherichia coli and Salmonella: Cellular and
1086 Molecular Biology. Washington, D.C.: 2256-2276. p. 2256-2276.

1087 Sanjuan R. 2010. Mutational fitness effects in RNA and single-stranded DNA viruses:
1088 common patterns revealed by site-directed mutagenesis studies. Philosophical
1089 transactions of the Royal Society of London. Series B, Biological sciences
1090 365:1975-1982. http://dx.doi.org/10.1098/rstb.2010.0063

1091 Sankar TS, Wastuwidyaningtyas BD, Dong Y, Lewis SA, Wang JD. 2016. The nature
1092 of mutations induced by replication-transcription collisions. Nature.
1093 http://dx.doi.org/10.1038/nature18316

1094 Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV,
1095 Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, et al. 2016. Local fitness
1096 landscape of the green fluorescent protein. Nature 533:397-401.
1097 http://dx.doi.org/10.1038/nature17995

1098 Seier T, Padgett DR, Zilberberg G, Sutera VA, Jr., Toha N, Lovett ST. 2011. Insights
1099 into mutagenesis using Escherichia coli chromosomal lacZ strains that enable
1100 detection of a wide spectrum of mutational events. Genetics 188:247-262.
1101 http://dx.doi.org/10.1534/genetics.111.127746

1102 Shewaramani S, Finn TJ, Leahy SC, Kassen R, Rainey PB, Moon CD. 2017.
1103 Anaerobically Grown Escherichia coli Has an Enhanced Mutation Rate and
1104 Distinct Mutational Spectra. PLoS genetics 13:e1006570.
1105 http://dx.doi.org/10.1371/journal.pgen.1006570

1106 Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C. 2005.
1107 Role of FRIGIDA and FLOWERING LOCUS C in determining variation in
1108 flowering time of Arabidopsis. Plant physiology 138:1163-1173.
1109 http://dx.doi.org/10.1104/pp.105.061309

1110 Silby MW, Cerdeno-Tarraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston
1111 GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SA, et al. 2009. Genomic and
1112 genetic analyses of diversity and plant interactions of Pseudomonas fluorescens.
1113 Genome biology 10:R51. http://dx.doi.org/10.1186/gb-2009-10-5-r51

1114 Simpson GG. 1949. The meaning of evolution : a study of the history of life and of its
1115 significance for man. New Haven: Yale University Press.

1116 Spiers AJ, Bohannon J, Gehrig SM, Rainey PB. 2003. Biofilm formation at the air-
1117 liquid interface by the Pseudomonas fluorescens SBW25 wrinkly spreader requires
1118 an acetylated form of cellulose. Molecular Microbiology 50:15-27.

1119 Spiers AJ, Kahn SG, Bohannon J, Travisano M, Rainey PB. 2002. Adaptive
1120 divergence in experimental populations of Pseudomonas fluorescens. I. Genetic
1121 and phenotypic bases of wrinkly spreader fitness. Genetics 161:33-46.

1122 Steinrueck M, Guet CC. 2017. Complex chromosomal neighborhood effects
1123 determine the adaptive potential of a gene under selection. Elife 6.
1124 http://dx.doi.org/10.7554/eLife.25100

1125 Stern DL. 2013. The genetic causes of convergent evolution. Nature reviews.
1126 Genetics 14:751-764. http://dx.doi.org/10.1038/nrg3483

1127 Stern DL, Orgogozo V. 2009. Is genetic evolution predictable? Science 323:746-751.
1128 http://dx.doi.org/10.1126/science.1158997

1129 Stoltzfus A, McCandlish DM. 2017. Mutational Biases Influence Parallel Adaptation.
1130 Molecular biology and evolution 34:2163-2172.
1131 http://dx.doi.org/10.1093/molbev/msx180

1132 Streisinger G, Owen J. 1985. Mechanisms of spontaneous and induced frameshift
1133 mutation in bacteriophage T4. Genetics 109:633-659.

1134 Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. 2012. Drift-barrier
1135 hypothesis and mutation-rate evolution. Proceedings of the National Academy of
1136 Sciences of the United States of America 109:18488-18492.
1137 http://dx.doi.org/10.1073/pnas.1216223109

1138 Tautz D, Renz M. 1984. Simple sequences are ubiquitous repetitive components of
1139 eukaryotic genomes. Nucleic Acids Res 12:4127-4138.

1140 Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD,
1141 Gaut BS. 2012. The molecular diversity of adaptive convergence. Science
1142 335:457-461. http://dx.doi.org/10.1126/science.1212986

1143 Ungerer MC, Linder CR, Rieseberg LH. 2003. Effects of genetic background on
1144 response to selection in experimental populations of Arabidopsis thaliana. Genetics
1145 163:277-286.

1146 Viswanathan M, Lacirignola JJ, Hurley RL, Lovett ST. 2000. A novel mutational
1147 hotspot in a natural quasipalindrome in Escherichia coli. J Mol Biol 302:553-564.
1148 http://dx.doi.org/10.1006/jmbi.2000.4088

1149 Wang AD, Sharp NP, Agrawal AF. 2014. Sensitivity of the distribution of mutational
1150 fitness effects to environment, genetic background, and adaptedness: a case study
1151 with Drosophila. Evolution 68:840-853. http://dx.doi.org/10.1111/evo.12309

1152 Wei W, Ning LW, Ye YN, Li SJ, Zhou HQ, Huang J, Guo FB. 2014. SMAL: A
1153 Resource of Spontaneous Mutation Accumulation Lines. Molecular biology and
1154 evolution 31:1302-1308. http://dx.doi.org/10.1093/molbev/msu073

1155 Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE. 2006. Tests of parallel
1156 molecular evolution in a long-term experiment with Escherichia coli. Proceedings
1157 of the National Academy of Sciences of the United States of America 103:9107-
1158 9112. http://dx.doi.org/10.1073/pnas.0602917103

1159 Xu M, Yang X, Yang XA, Zhou L, Liu TZ, Fan Z, Jiang T. 2016. Structural insights
1160 into the regulatory mechanism of the Pseudomonas aeruginosa YfiBNR system.
1161 Protein Cell 7:403-416. http://dx.doi.org/10.1007/s13238-016-0264-7

1162 Zhao HL, Xia ZK, Zhang FZ, Ye YN, Guo FB. 2015. Multiple Factors Drive
1163 Replicating Strand Composition Bias in Bacterial Genomes. Int J Mol Sci
1164 16:23111-23126. http://dx.doi.org/10.3390/ijms160923111

1165 Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. 2012. Parallel
1166 molecular evolution in an herbivore community. Science 337:1634-1637.
1167 http://dx.doi.org/10.1126/science.1226630

1168

Figure supplements

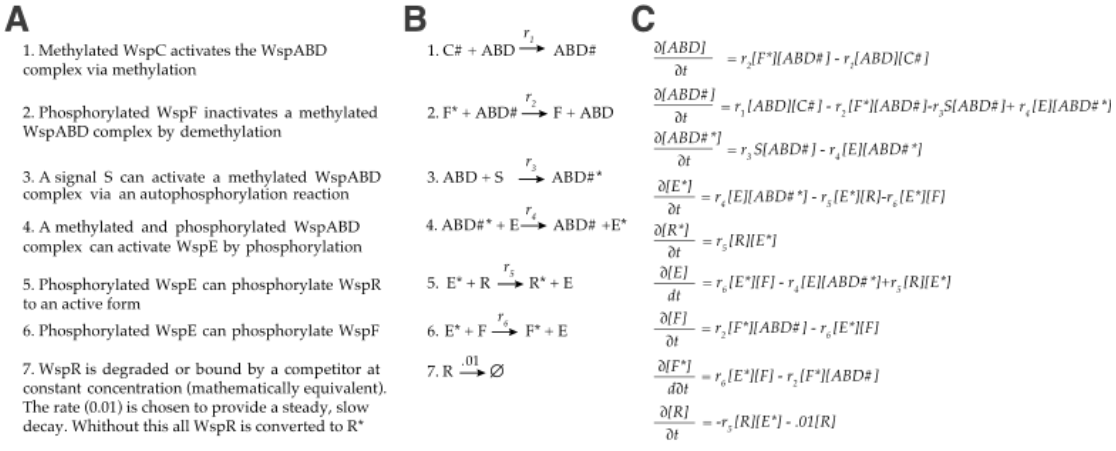


Figure 3 - figure supplement 1. Wsp model (A) Description of functional interactions **(B)** Description of molecular reaction **(C)** Differential equations describing the dynamics of the Wsp pathway. The activity of WspA is modulated by methylation, where it is activated by the CheR-like methyltransferase 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.

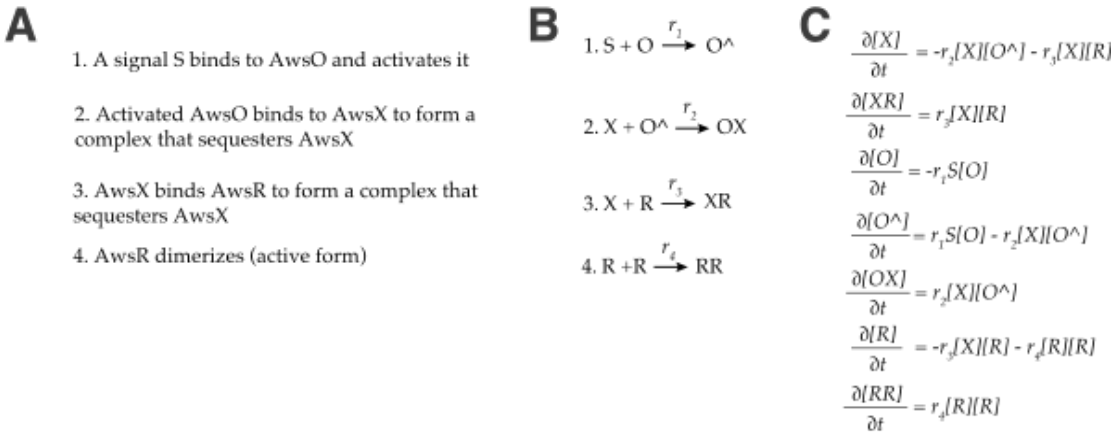
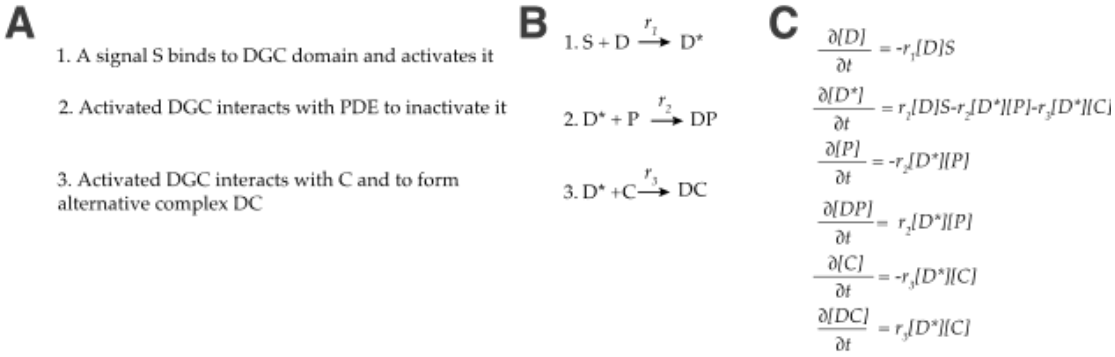


Figure 3 - figure supplement 2. Aws model (A) Description of functional interactions **(B)** Description of molecular reaction **(C)** Differential equations describing the dynamics of the Aws pathway. Release of AwsX mediated repression results in a conformational shift that in the model is represented as formation of an active dimer.

1187



1188

1189 **Figure 3 - figure supplement 3. Mws model (A)** Description of functional

1190 interactions **(B)** Description of molecular reaction **(C)** Differential equations

1191 describing the dynamics of the Aws pathway

1192

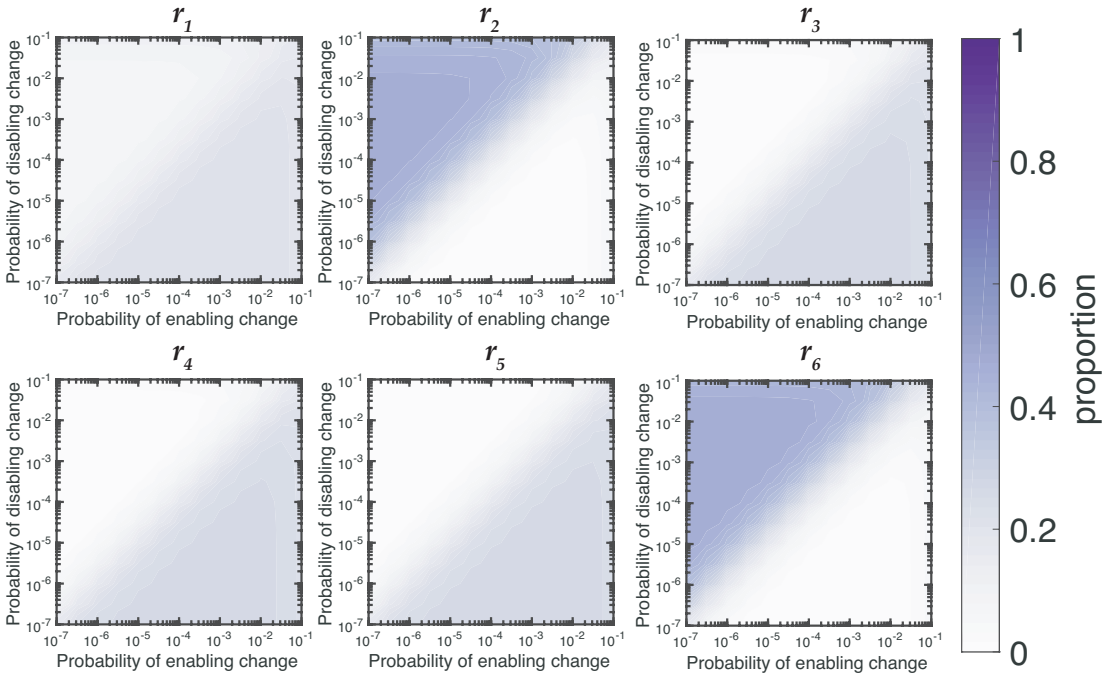


Figure 4 - figure supplement 1. Relative contributions of reaction rates for Wsp

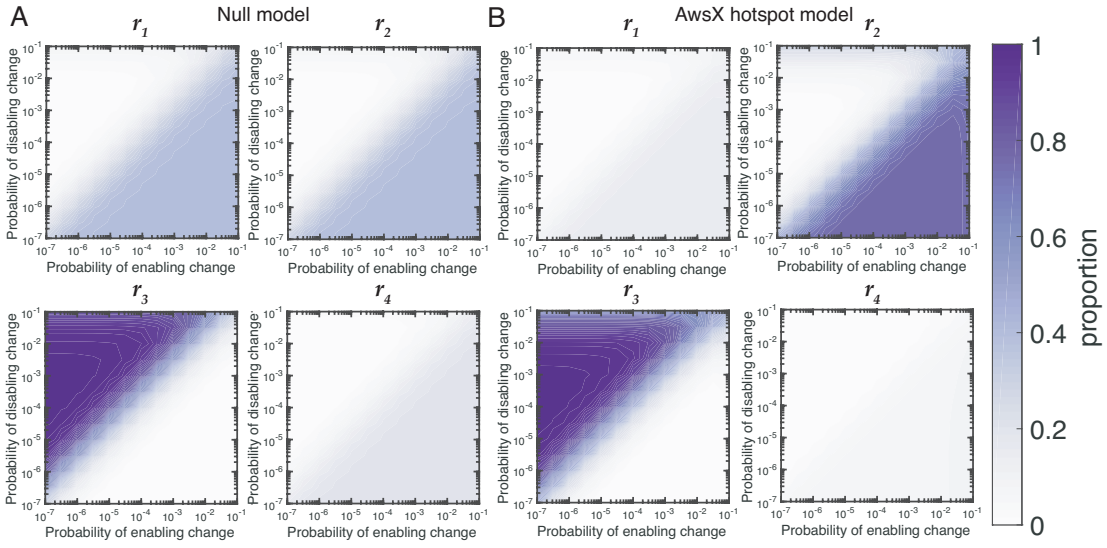


Figure 4 - figure supplement 2. Relative contributions of reaction rates for Aws.

Effect of the mutational hotspot on the relative contribution of each reaction in Aws to the probability of generating a wrinkly spreader **(A)** Null model. **(B)** Hot spot in AwsX increase rate mutation rate five times. Because the mutational hotspot increases the likelihood that AwsX's reactions will be altered, r_2 and r_3 have an increased probability of enabling or disabling change. As a consequence, r_2 shows the biggest change as it contributes up to 70% (as opposed to 35% without the mutational hotspot) of the probability that Aws is used to generate a WS in the area where

enabling change is more likely than disabling change. Furthermore, in the other area of parameter space where disabling change is more likely, r_3 contributes up to 90% of the probability that Aws is used with or without the hotspot.

Supplementary files

Figure S1

Figure 5 – source data 1. Table of all WS mutations in Wsp, Aws and Mws.

ODE model and code

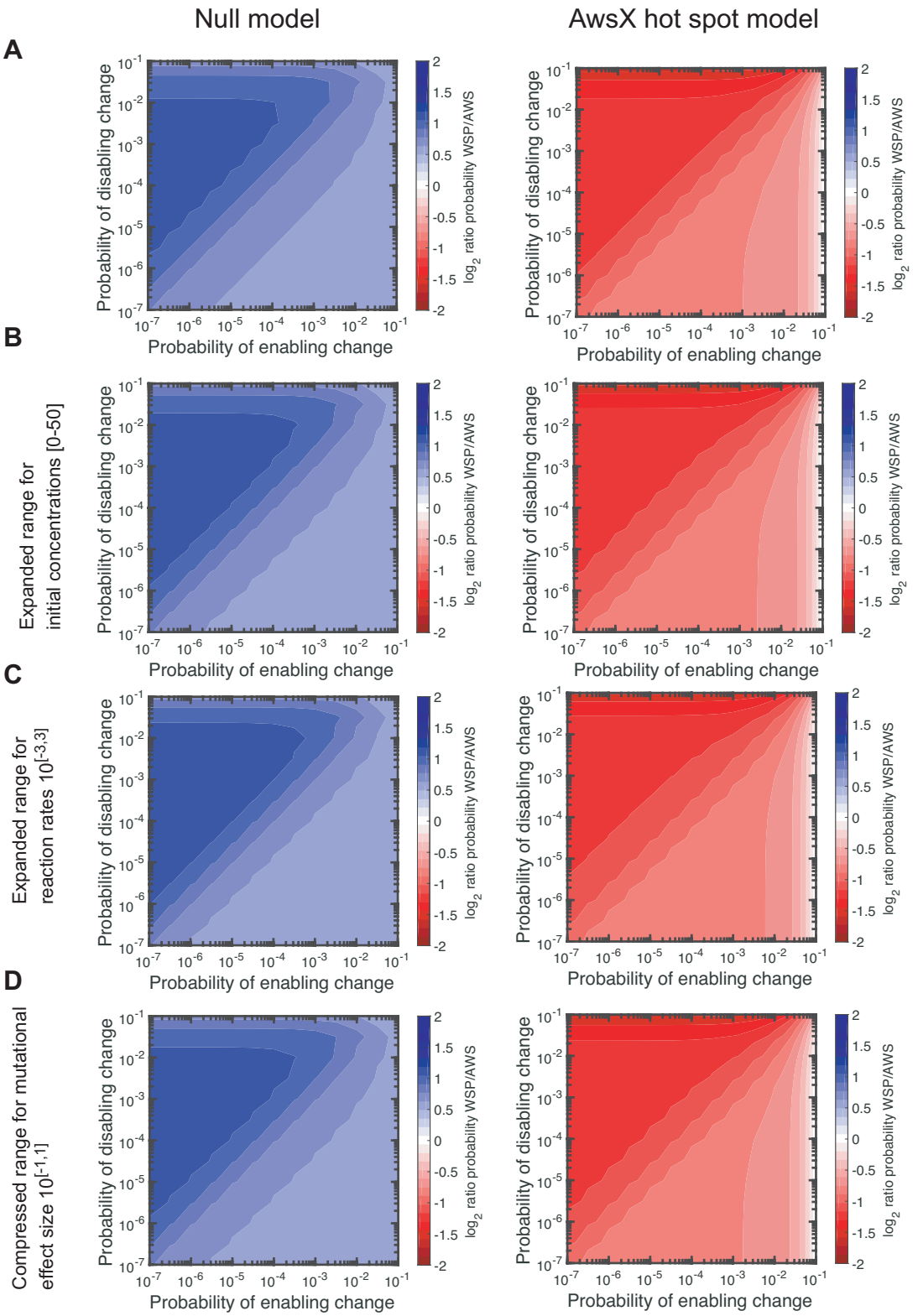


Figure S1. **Parameter sensitivity analysis.** To assess the effect of the chosen parameter (**A**) ranges on our results, we redid our sampling procedure for WSP for three different parameter regimes: (**B**) an expanded range for initial concentrations [0-50], (**C**) an expanded range for reaction rates $10^{[-3,3]}$, (**D**) a compressed range for mutational effect size $10^{[-1,1]}$. We found that our qualitative results are robust to these changes**.**

Figure 5 - Source data - Wsp mutations

Genome position	Type	Change	Gene positio	Gene locus	Gene symbol	Effect	Comment	Proposed molecular effect
1353597 - 1353602	Deletion	del 6 bp	1328 - 1333	PFLU1219	*wspA*	In frame del G443 - M444)	No homology	
1353112 - 1353195	Deletion	del 84 bp	843 - 926	PFLU1219	*wspA*	In frame del A281-A308	6 bp GGCCAC homology	methylation site
1353112 - 1353195	Deletion	del 84 bp	843 - 926	PFLU1219	*wspA*	In frame del A281-A308	6 bp GGCCAC homology	methylation site
1353140 - 1353160	Deletion	del 21 bp	878 - 898	PFLU1219	*wspA*	In frame del T293 - E299	5 bp ACTGA homology	methylation site
1353140 - 1353160	Deletion	del 21 bp	878 - 898	PFLU1219	*wspA*	In frame del T293 - E299	5 bp ACTGA homology	methylation site
1353423	Transversion	A->C	1154	PFLU1219	*wspA*	N385T		trimer-of-dimer association
1353451	Transversion	A->C	1182	PFLU1219	*wspA*	E394D		trimer-of-dimer association
1353451	Transversion	A->C	1182	PFLU1219	*wspA*	E394D		trimer-of-dimer association
1353411	Transition	C->T	1142	PFLU1219	*wspA*	A381V		trimer-of-dimer association
1353411	Transition	C->T	1142	PFLU1219	*wspA*	A381V		trimer-of-dimer association
1353528	Transition	C->T	1259	PFLU1219	*wspA*	A420V		possible trimer-of-dimer association
1353528	Transition	C->T	1259	PFLU1219	*wspA*	A420V		possible trimer-of-dimer association
1353323	Transition	G->A	1054	PFLU1219	*wspA*	A352T		possible trimer-of-dimer association
1353323	Transition	G->A	1054	PFLU1219	*wspA*	A352T		possible trimer-of-dimer association
1353408	Transversion	T->G	1139	PFLU1219	*wspA*	V380G		trimer-of-dimer association
1355373 - 1355707	Deletion	del 309 bp	973 - 51	PFLU1221	*wspC/D*	WspC M1-A324 fused to WspD D18-S232	9 bp ACCCTGGCC homology	increase wspC activity
1358267	Transversion	A->C	1916	PFLU1223	*wspE*	D639A		disrupt phoshorylation site of wspF
1358267	Transversion	A->C	1916	PFLU1223	*wspE*	D639A		disrupt phoshorylation site of wspF
1358267	Transversion	A->C	1916	PFLU1223	*wspE*	D639A		disrupt phoshorylation site of wspF
1358267	Transversion	A->C	1916	PFLU1223	*wspE*	D639A		disrupt phoshorylation site of wspF
1358267	Transversion	A->C	1916	PFLU1223	*wspE*	D639A		disrupt phoshorylation site of wspF
1358551	Transition	A->G	2200	PFLU1223	*wspE*	K734E		disrupt phoshorylation site of wspF
1358553	Transversion	A->C	2202	PFLU1223	*wspE*	K734N		disrupt phoshorylation site of wspF
1358553	Transversion	A->C	2202	PFLU1223	*wspE*	K734N		disrupt phoshorylation site of wspF
1358553	Transversion	A->C	2202	PFLU1223	*wspE*	K734N		disrupt phoshorylation site of wspF
1358270	Transition	C->T	1919	PFLU1223	*wspE*	S640L		disrupt phoshorylation site of wspF
1358266	Transversion	G->T	1915	PFLU1223	*wspE*	D639Y		disrupt phoshorylation site of wspF
1358279	Transversion	T->A	1928	PFLU1223	*wspE*	V643E		disrupt phoshorylation site of wspF
1358279	Transversion	T->A	1928	PFLU1223	*wspE*	V643E		disrupt phoshorylation site of wspF
1358766 - 1358780	Deletion	del 15 bp	151-165	PFLU1224	*wspF*	In frame del L51- I55	10 bp GGACCTGATC homology	disrupt demethylase activity
1358766 - 1358780	Deletion	del 15 bp	151-165	PFLU1224	*wspF*	In frame del L51- I55	10 bp GGACCTGATC homology	disrupt demethylase activity
1358810 -1358935	Deletion	del 126 bp	195 - 320	PFLU1224	*wspF*	In frame del R66 -L107	No homology	disrupt demethylase activity
1359285 - 1359289	Deletion	del 5 bp	670 - 674	PFLU1224	*wspF*	Frame shift after L223	3 bp GGC homology	disrupt demethylase activity
1359292 - 1359441	Deletion	del 150 bp	677 - 826	PFLU1224	*wspF*	In frame del T226-G275	3 bp ACC homology	disrupt demethylase activity
1359505	Transition	A->G	890	PFLU1224	*wspF*	Q297R		disrupt demethylase activity

1358755	Transition	C->T	140	PFLU1224	*wspF*	P47L		disrupt demethylase activity
1359091	Transition	C->T	476	PFLU1224	*wspF*	S159L		disrupt demethylase activity
1359171	Transition	C->T	556	PFLU1224	*wspF*	H186Y		disrupt demethylase activity
1359538	Transition	C->T	923	PFLU1224	*wspF*	P308L		disrupt demethylase activity
1359423	Transversion	G->C	808	PFLU1224	*wspF*	G270R		disrupt demethylase activity
1359604	Transversion	T->A	989	PFLU1224	*wspF*	L331*		disrupt demethylase activity
1360156 - 1360167	Deletion	del 12 bp	481 - 492	PFLU1225	*wspR*	In frame del N161 - L164	No homology	disrupt requirement for phosphorylation
1360099	Transversion	C->A	424	PFLU1225	*wspR*	Q142K		disrupt requirement for phosphorylation
1360102	Transition	C->T	427	PFLU1225	*wspR*	R143W		disrupt requirement for phosphorylation
1360102	Transition	C->T	427	PFLU1225	*wspR*	R143W		disrupt requirement for phosphorylation
1360107	Transversion	T->G	432	PFLU1225	*wspR*	D144E		disrupt requirement for phosphorylation

Figure 5 - Source data - Aws mutations

Genome position	Type	Change	Gene positio	Gene locus	Gene symbol	Effect	Comment	Proposed molecular effect
5705729	Transversion	A->C	101	PFLU5209	*awsO*	Q34P		change binding with awsX
5705711	Transversion	A->C	119	PFLU5209	*awsO*	Q40P		change binding with awsX
5706522	Transition	C->T	575	PFLU5210	*awsR*	A192V		HAMP linker
5706522	Transition	C->T	575	PFLU5210	*awsR*	A192V		HAMP linker
5706909	Transition	C->T	188	PFLU5210	*awsR*	A63V		interaction awsX
5706441	Transversion	A->C	656	PFLU5210	*awsR*	D219A		HAMP linker
5706937	Transition	C->T	160	PFLU5210	*awsR*	R54C		interaction awsX
5707018	Transversion	A->C	79	PFLU5210	*awsR*	T27P		disrupt TM helix
5707018	Transversion	A->C	79	PFLU5210	*awsR*	T27P		disrupt TM helix
5707018	Transversion	A->C	79	PFLU5210	*awsR*	T27P		disrupt TM helix
5707018	Transversion	A->C	79	PFLU5210	*awsR*	T27P		disrupt TM helix
5707018	Transversion	A->C	79	PFLU5210	*awsR*	T27P		disrupt TM helix
5707018	Transversion	A->C	79	PFLU5210	*awsR*	T27P		disrupt TM helix
5707018	Transversion	A->C	79	PFLU5210	*awsR*	T27P		disrupt TM helix
5707018	Transversion	A->C	79	PFLU5210	*awsR*	T27P		disrupt TM helix
5707018	Transversion	A->C	79	PFLU5210	*awsR*	T27P		disrupt TM helix
5707574	Transition	C->T	92	PFLU5211	*awsX*	A31V		disrupt awsX function
5707528-5707566	Deletion	del 39 bp	100-138	PFLU5211	*awsX*	del P34-A46	10 bp homology CGCCCAGGCG	disrupt awsX function
5707528-5707566	Deletion	del 39 bp	100-138	PFLU5211	*awsX*	del P34-A46	10 bp homology CGCCCAGGCG	disrupt awsX function
5707460	Transversion	T->G	206	PFLU5211	*awsX*	L69R		disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function

5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707405-5707437	Deletion	del 33 bp	229-261	PFLU5211	*awsX*	del Y77-Q87	6 bp homology ACCCAG	disrupt awsX function
5707262	Transversion	T->G	404	PFLU5211	*awsX*	L135R		disrupt awsX function

Figure 5 - Source data - Mws mutations

Genome position	Type	Change	Gene positio	Gene locus	Gene symbol	Effect	Comment	Proposed molecular effect
5857733-5857741	Deletion	del 9 bp	3071-3079	PFLU5329	*mwsR*	del R1024 - E1026	8 bp homology CCTGGAGC	Disrupt interaction DGC and EAL domain
5857733-5857741	Deletion	del 9 bp	3071-3079	PFLU5329	*mwsR*	del R1024 - E1026	8 bp homology CCTGGAGC	Disrupt interaction DGC and EAL domain
5857733-5857741	Deletion	del 9 bp	3071-3079	PFLU5329	*mwsR*	del R1024 - E1026	8 bp homology CCTGGAGC	Disrupt interaction DGC and EAL domain
5857733-5857741	Deletion	del 9 bp	3071-3079	PFLU5329	*mwsR*	del R1024 - E1026	8 bp homology CCTGGAGC	Disrupt interaction DGC and EAL domain
5857733-5857741	Deletion	del 9 bp	3071-3079	PFLU5329	*mwsR*	del R1024 - E1026	8 bp homology CCTGGAGC	Disrupt interaction DGC and EAL domain
5857733-5857741	Deletion	del 9 bp	3071-3079	PFLU5329	*mwsR*	del R1024 - E1026	8 bp homology CCTGGAGC	Disrupt interaction DGC and EAL domain
5857733-5857741	Deletion	del 9 bp	3071-3079	PFLU5329	*mwsR*	del R1024 - E1026	8 bp homology CCTGGAGC	Disrupt interaction DGC and EAL domain
5857733-5857741	Deletion	del 9 bp	3071-3079	PFLU5329	*mwsR*	del R1024 - E1026	8 bp homology CCTGGAGC	Disrupt interaction DGC and EAL domain
5857733-5857741	Deletion	del 9 bp	3071-3079	PFLU5329	*mwsR*	del R1024 - E1026	8 bp homology CCTGGAGC	Disrupt interaction DGC and EAL domain
5857733-5857741	Deletion	del 9 bp	3071-3079	PFLU5329	*mwsR*	del R1024 - E1026	8 bp homology CCTGGAGC	Disrupt interaction DGC and EAL domain
5857733-5857741	Deletion	del 9 bp	3071-3079	PFLU5329	*mwsR*	del R1024 - E1026	8 bp homology CCTGGAGC	Disrupt interaction DGC and EAL domain
A5857192C	Transversion	A->C	A2530C	PFLU5329	*mwsR*	I844L		
C5857598/ C5857599	Insertion	duplication 2	2937-2959	PFLU5329	*mwsR*	dup I978-G985		Disrupt interaction DGC and EAL domain
G5857440A	Transition	G->A	G2778A	PFLU5329	*mwsR*	M926I		Disrupt interaction M992
G5857440A	Transition	G->A	G2778A	PFLU5329	*mwsR*	M926I		Disrupt interaction M926
G5857440A	Transition	G->A	G2778A	PFLU5329	*mwsR*	M926I		Disrupt interaction M926
G5857638A	Transition	G->A	G2976A	PFLU5329	*mwsR*	M992I		Disrupt interaction M926
G5857909A	Transition	G->A	G3247A	PFLU5329	*mwsR*	E1083K		Disrupt interaction DGC and EAL domain

Differential equations for WSP, AWS, & MWS pathways

WSP

$$
\begin{aligned}
\frac{d[ABD]}{dt} &= r_2F^*[ABDm] - r_1[ABD][Cm] \\
\frac{d[ABDm]}{dt} &= r_1[ABD][Cm] - r_2F^*[ABDm] - r_3S[ABDm] + r_4E[ABDm^*] \\
\frac{d[ABDm^*]}{dt} &= -r_4E[ABDm^*] + r_3S[ABDm] \\
\frac{dE^*}{dt} &= r_4E[ABDm^*] - r_5E^*R - r_6E^*F \\
\frac{dR^*}{dt} &= r_5RE^* \\
\frac{dE}{dt} &= r_6E^*F - r_4E[ABDm^*] + r_5RE^* \\
\frac{dF}{dt} &= -r_6E^*F + r_2F^*[ABDm] \\
\frac{dF^*}{dt} &= r_6E^*F - r_2F^*[ABDm] \\
\frac{dR}{dt} &= -r_5RE^* - .01R
\end{aligned}
$$

AWS

$$
\begin{aligned}
\frac{dX}{dt} &= -r_2XO^* - r_3XR \\
\frac{d[XR]}{dt} &= r_3XR \\
\frac{dO}{dt} &= -r_1SO \\
\frac{dO^*}{dt} &= r_1SO - r_2XO^* \\
\frac{d[OX]}{dt} &= r_2O^*X \\
\frac{dR}{dt} &= -r_3XR - r4RR \\
\frac{d[RR]}{dt} &= r_4RR
\end{aligned}
$$

MWS

$$
\begin{aligned}
\frac{dG}{dt} &= -r_1 GS \\
\frac{dG^*}{dt} &= r_1 GS - r_2 G^* E - r_3 G^* C \\
\frac{dE}{dt} &= -r_2 G^* E \\
\frac{d[GE]}{dt} &= r_2 G^* E \\
\frac{dC}{dt} &= -r_3 G^* C \\
\frac{d[GC]}{dt} &= r3 G^* C
\end{aligned}
$$

Julia code for WSP, AWS, and MWS differential equations

The following three functions implement the differential equation model (ODE model) for the WSP, AWS, and MWS pathways.

```
using ODE,StatsBase,MAT,HDF5,JLD

# code for WSP differential equations
function lindodeWSP(t,y)
   # convert y to reactants for ease of reading
   ABD=y[1];
   ABDm=y[2];
   ABDmp=y[3];
   Ep=y[4];
   Rp=y[5];
   E=y[6];
   F=y[7];
   Fp=y[8];
   R=y[9];
   # pull reaction rates from rs variable
   r1=rs[1];r2=rs[2];r3=rs[3];r4=rs[4];r5=rs[5];r6=rs[6];
   # compute derivatives, i.e. y'
   yp=zeros(size(y));
   yp[1]=r2*Fp*ABDm-r1*ABD*Cm; # dABD/dt
   yp[2]=r1*ABD*Cm-r2*Fp*ABDm-r3*S*ABDm+r4*E*ABDmp; # dABDm/dt
   yp[3]=-r4*E*ABDmp + r3*S*ABDm ; # dABDmp/dt
   yp[4]=r4*E*ABDmp - r5*Ep*R -r6*Ep*F ; # dEp/dt
   yp[5]=r5*R*Ep ; # dRp/dt
   yp[6]=r6*Ep*F-r4*E*ABDmp+r5*R*Ep ; # dE/dt
   yp[7]=-r6*Ep*F +r2*Fp*ABDm ; # dF/dt
   yp[8]=r6*Ep*F -r2*Fp*ABDm ; # dFp/dt
   yp[9]=-r5*R*Ep-.01*R; # dR/dt
   return yp
end

# code for AWS differential equations
```

```
function lindodeAWS(t,y)
   # convert y to reactants for ease of reading
   X=y[1];
   XR=y[2];
   O=y[3];
   Op=y[4];
   OX=y[5];
   R=y[6];
   RR=y[7];
   # pull reaction rates from rs variable
   r1=rs[1];r2=rs[2];r3=rs[3];r4=rs[4];
   # compute derivatives, i.e. y'
   yp=zeros(size(y));
   yp[1]=-r2*X*Op-r3*X*R; # dX/dt
   yp[2]=r3*X*R; # dXR/dt
   yp[3]=-r1*S*O; # dO/dt
   yp[4]=r1*S*O-r2*X*Op; # dOp/dt
   yp[5]=r2*Op*X; # dOX/dt
   yp[6]=-r3*X*R-r4*R*R; # dR/dt
   yp[7]=r4*R*R; # dRR/dt
   return yp;
end

# code for MWS differential equations
function lindodeMWS(t,y)
   # convert y to reactants for ease of reading
   G=y[1];
   Gp=y[2];
   E=y[3];
   GE=y[4];
   C=y[5];
   GC=y[6];
   # pull reaction rates from rs variable
   r1=rs[1];r2=rs[2];r3=rs[3];
   # compute derivatives, i.e. y'
   yp=zeros(size(y));
   yp[1]=-r1*G*S; # dG/dt
   yp[2]=r1*G*S-r2*Gp*E-r3*Gp*C;# dGp/dt
   yp[3]= -r2*Gp*E;# dE/dt
   yp[4]= r2*Gp*E;# dGE/dt
   yp[5]=-r3*Gp*C;# dC/dt
   yp[6]=r3*Gp*C; # dGC/dt
   return yp;
end
```

Julia code for running differential equation solvers

We include the code for implementing our Bayesian sampling method. The colored sections correspond to statements that make it specific to WSP (blue), AWS (red), or MWS (green). It was run in julia version 0.4.3.

```
# Variables for reactions
numrxns=6; # number of reaction rates for WSP
```

```
numrxns=4; # number of reaction rates for AWS
numrxns=3; # number of reaction rates for MWS
rs=rand(numrxns); # establish variable scope, will be reaction rates later
rs_save=copy(rs); # establish variable scope, will be a saved version of reaction rates later
totnumruns=3.^length(rs); # all possible combinations for reaction rates (down,nothing, up)
v=zeros(length(rs)); # establish variable scope (used to alter reaction rates)
S=0;Cm=0; # initialize constants used in differential equations
indexWS=5; # reactant corresponding to WS in WSP
indexWS=7; # reactant corresponding to WS in AWS
indexWS=6; # reactant corresponding to WS in MWS

# Variables for running ODE solver
testnums=1000; # number of runs
yorig=zeros(testnums); # storage for baseline WS production
yout=0; # establish variable scope
tf=1.0; # establish variable scope
tftimes=1.0; # establish variable scope

numreactants=9; # number of reactants
numreactants=7; # number of reactants
numreactants=6; # number of reactants
init=10*rand(numreactants); # establish variable scope


# Variables for storing data
res=-1*ones(totnumruns,testnums); # storage for altered WS production
numfinished=1; # counter for runs completed

# Code for Bayesian sampling method
while numfinished<=testnums
   done=0;
   try
      println(numfinished) # keeps track of how many sims have been done
      # Sample concentrations and rates to establish a baseline amount of WS production
      rs=10.^(4*rand(numrxns)-2); # sample reaction rates from [.01,100]
      rs_save=copy(rs); # saved copy as a reference when altering later
      init=10*rand(numreactants); # sample initial concentrations for reactants from [0,10]
      S=10*rand(); # sample initial concentration for constant reactant of signal (S) from [0,10]
      Cm=10*rand(); # sample initial concentration for constant reactant Cm from [0,10]

      # ODE solver for baseline
      tf=1.0; # initial time for ode solver
      dst=100; # initial distance, used to determine solution converged
      tol=10^(-8.0); # tolerance for ODE solver
      while dst>tol
         tout, yout = ode45(lindodeWSP, init, [0.0 ,tf]);
         tout, yout = ode45(lindodeAWS, init, [0.0 ,tf]);
         tout, yout = ode45(lindodeMWS, init, [0.0 ,tf]);
         dst=sum((yout[end-1]-yout[end]).^2); # Euclidean distance in final step of solution
         yorig[numfinished]=yout[end][indexWS]; # reactant corresponding to WS
         tftimes=tout[end]; # final time
         tf*=2;
      end
      done=1; # successful completion of try loop
   end
   if done==1 # baseline WS production is established, now sample changes/mutations
      i0=1; # counter for completed changes
      cct=0; # counter for total number of attempts
```

```
      while i0<=totnumruns;
         cct+=1;
         println([numfinished cct]) # report status for tracking progress
         # Alter reaction rates
         num=i0-1; # used for determining which rates change down/none/up
            for i1=length(rs)-1:-1:0;
                  v[i1+1]=floor(num/(3^i1)); # v is num into base 3 number
                  num=num-v[i1+1]*3^i1;
            end
         v+=1;
         facs=[10.^(-2*rand()), 1, 10.^(2*rand())]; # factors to alter rxn rates [.01,1] down, 1 none, [1,100] up
         for i1=1:length(rs)
            rs[i1]=facs[v[i1]]*rs_save[i1]; # alter reaction rates
         end
         try
            tout, yout = ode45(lindodeWSP, init, [0.0,tftimes]);
            tout, yout = ode45(lindodeAWS, init, [0.0,tftimes]);
            tout, yout = ode45(lindodeMWS, init, [0.0,tftimes]);
            if abs(tout[end]-tftimes)<.01 # ODE solver finished
                     res[i0,numfinished]=yout[end][indexWS]; # store amount of WS produced
               i0+=1
            end
         end
         if cct>10000
            i0=2*totnumruns; # baseline and sampling occurred in space with poorly conditioned ODEs, try again
         end
      end
      if i0<2*totnumruns # successful
         numfinished+=1;
         # Save data to a file for checking in MATLAB
         file=matopen("pathway_results_temp.mat","w")
         write(file,"res",res); altered WS production
         write(file,"yorig",yorig); baseline WS production
         close(file);
      end
      end
end

# Save data to a file for processing in MATLAB
file=matopen("pathway_results_complete_WSP.mat","w")
file=matopen("pathway_results_complete_AWS.mat","w")
file=matopen("pathway_results_complete_MWS.mat","w")
write(file,"res",res); altered WS production
write(file,"yorig",yorig); baseline WS production
close(file);
```

MATLAB code for interpreting saved data WSP vs MWS

This code shows how the data from the julia code is analyzed and transformed into the contour plots shown in the paper.

```
% create record of how parameters change down, none, up for WSP
rs1=rand(1,6);
```

```
totnumruns=3.^length(rs1);
paramsWSP=zeros(totnumruns,length(rs1));
v=zeros(size(rs1));
i0=1;
while i0<=totnumruns;
    num=i0-1;
    for i1=length(rs1)-1:-1:0;
            v(i1+1)=floor(num/(3^i1));
            num=num-v(i1+1)*3^i1;
    end
    v=v+1;
    paramsWSP(i0,:)=v;
    i0=i0+1;
end
% create record of how parameters change down, none, up for MWS
rs1=rand(1,3);
totnumruns=3.^length(rs1);
paramsMWS=zeros(totnumruns,length(rs1));
v=zeros(size(rs1));
i0=1;
while i0<=totnumruns;
    num=i0-1;
    for i1=length(rs1)-1:-1:0;
            v(i1+1)=floor(num/(3^i1));
            num=num-v(i1+1)*3^i1;
    end
    v=v+1;
    paramsMWS(i0,:)=v;
    i0=i0+1;
end

% Load data
load pathway_results_complete_MWS.mat
resMWS=res;
yorigMWS=yorig';
clear res yorig
load pathway_results_complete_WSP.mat
resWSP=res;
yorigWSP=yorig';

% Variables and data storage to compare likelihood of pathways
numsampWSP=size(resWSP,2); % in case want to use fewer samples
numsampMWS=size(resMWS,2); % in case want to use fewer samples

perange=10.^[-7:.5:-1]; % range for probability enabling mutations
pdrange=10.^[-7:.5:-1]; % range for probability disabling mutations
pemat=zeros(length(perange),length(pdrange)); % matrix for plotting data and reference
pdmat=zeros(size(pemat)); % matrix for plotting data and reference
psummatWSP=zeros(size(pemat)); % matrix for probability WSP used
psummatMWS=zeros(size(pemat)); % matrix for probability MWS used
tol=0;
% Code to compare likelihood of pathways
for i0=1:length(perange)
    for j0=1:length(pdrange)
        % Retrieve probabilities
        pe=perange(i0);
        pd=pdrange(j0);
        pemat(i0,j0)=pe;
```

```
        pdmat(i0,j0)=pd;
        % WSP computation
        pmatforr=ones(6,1)*[pd 1-pe-pd pe];
        psum=0; % total sum of (prob of rxn changes) X (number of times WS produced)
        for i1=1:size(resWSP,1)
            probevent=1; % initialize, probability to get combination of down, none, up for rxns
            for j1=1:6;
                probevent=probevent*pmatforr(j1,paramsWSP(i1,j1)); % multiply by prob of each change
                    end
            psum=psum+probevent*sum(resWSP(i1,1:numsampWSP)>yorigWSP(1:numsampWSP)+tol)/numsampWSP;
        end
        psummatWSP(i0,j0)=psum;

        % MWS computation
        pmatforr=ones(3,1)*[pd 1-pe-pd pe];
        psum=0; % total sum of (prob of rxn changes) X (number of times WS produced)
        for i1=1:size(resMWS,1)
            probevent=1; % initialize, probability to get combination of down, none, up for rxns
            for j1=1:3;
                probevent=probevent*pmatforr(j1,paramsMWS(i1,j1)); % multiply by prob of each change
            end
            psum=psum+probevent*sum(resMWS(i1,1:numsampMWS)>yorigMWS(1:numsampMWS)+tol)/numsampMWS;
        end
        psummatMWS(i0,j0)=psum;
    end
end

% Plot data
close all
figure
contourf(pemat,pdmat,log2(psummatWSP./psummatMWS),400,'LineStyle','None')
set(gca,'xScale','log','yScale','log','TickLength',[.025 .025],'LineWidth',3);
set(gca,'FontSize',18,'xTick',10.^[-7:1:-1],'yTick',10.^[-7:1:-1]);
xlabel('Probability of enabling change','FontSize',24);
ylabel('Probability of disabling change','FontSize',24);
c=colorbar('FontSize',18);
c.Label.String='log_2 ratio probability WSP/MWS';
colormap jet;
axis square
eval(['print -f1 -depsc -r300 WSP_vs_MWS_contour.eps']);
```

MATLAB code for interpreting saved data WSP vs AWS

```
% create record of how parameters change down, none, up for WSP
rs1=rand(1,6);
totnumruns=3.^length(rs1);
paramsWSP=zeros(totnumruns,length(rs1));
v=zeros(size(rs1));
i0=1;
while i0<=totnumruns;
    num=i0-1;
    for i1=length(rs1)-1:-1:0;
            v(i1+1)=floor(num/(3^i1));
```

```
            num=num-v(i1+1)*3^i1;
    end
    v=v+1;
    paramsWSP(i0,:)=v;
    i0=i0+1;
end
% create record of how parameters change down, none, up for AWS
rs1=rand(1,4);
totnumruns=3.^length(rs1);
paramsAWS=zeros(totnumruns,length(rs1));
v=zeros(size(rs1));
i0=1;
while i0<=totnumruns;
    num=i0-1;
    for i1=length(rs1)-1:-1:0;
            v(i1+1)=floor(num/(3^i1));
            num=num-v(i1+1)*3^i1;
    end
    v=v+1;
    paramsAWS(i0,:)=v;
    i0=i0+1;
end

% Load data
load pathway_results_complete_AWS.mat
resAWS=res;
yorigAWS=yorig';
clear res yorig
load pathway_results_complete_WSP.mat
resWSP=res;
yorigWSP=yorig';

% Variables and data storage to compare likelihood of pathways
numsampWSP=size(resWSP,2); % in case want to use fewer samples
numsampAWS=size(resAWS,2); % in case want to use fewer samples

perange=10.^[-7:.5:-1]; % range for probability enabling mutations
pdrange=10.^[-7:.5:-1]; % range for probability disabling mutations
pemat=zeros(length(perange),length(pdrange)); % matrix for plotting data and reference
pdmat=zeros(size(pemat)); % matrix for plotting data and reference
psummatWSP=zeros(size(pemat)); % matrix for probability WSP used
psummatAWS=zeros(size(pemat)); % matrix for probability MWS used
tol=0;
fac=5; % factor increase of mutation because of hotspot

% Code to compare likelihood of pathways
for i0=1:length(perange)
    for j0=1:length(pdrange)
        % Retrieve probabilities
        pe=perange(i0);
        pd=pdrange(j0);
        pemat(i0,j0)=pe;
        pdmat(i0,j0)=pd;
        % WSP computation
        pmatforr=ones(6,1)*[pd 1-pe-pd pe];
        psum=0; % total sum of (prob of rxn changes) X (number of times WS produced)
        for i1=1:size(resWSP,1)
            probevent=1; % initialize, probability to get combination of down, none, up for rxns
```

```
            for j1=1:6;
                probevent=probevent*pmatforr(j1,paramsWSP(i1,j1)); % multiply by prob of each change
                    end
            psum=psum+probevent*sum(resWSP(i1,1:numsampWSP)>yorigWSP(1:numsampWSP)+tol)/numsampWSP;
        end
        psummatWSP(i0,j0)=psum;

        % AWS computation
        pmatforr=ones(4,1)*[pd 1-pe-pd pe];
        pmatforr(4,:)=[.5*pd 1-.5*pd-.5*pe .5*pe]; % because only reactant in dimerization
        pmatforr(3,:)=[fac*pd 1-fac*pd-fac*pe fac*pe]; % effect of hotspot
        pmatforr(2,:)=[fac*pd 1-fac*pd-fac*pe fac*pe]; % effect of hotspot
        psum=0; % total sum of (prob of rxn changes) X (number of times WS produced)
        for i1=1:size(resAWS,1)
            probevent=1; % initialize, probability to get combination of down, none, up for rxns
            for j1=1:4;
                probevent=probevent*pmatforr(j1,paramsAWS(i1,j1)); % multiply by prob of each change
            end
            psum=psum+probevent*sum(resAWS(i1,1:numsampAWS)>yorigAWS(1:numsampAWS)+tol)/numsampAWS;
        end
        psummatAWS(i0,j0)=psum;
    end
end

% Plot data
close all
figure
contourf(pemat,pdmat,log2(psummatWSP./psummatAWS),400,'LineStyle','None')
set(gca,'xScale','log','yScale','log','TickLength',[.025 .025],'LineWidth',3);
set(gca,'FontSize',18,'xTick',10.^[-7:1:-1],'yTick',10.^[-7:1:-1]);
xlabel('Probability of enabling change','FontSize',24);
ylabel('Probability of disabling change','FontSize',24);
c=colorbar('FontSize',18);
c.Label.String='log_2 ratio probability WSP/AWS';
colormap jet;
axis square
eval(['print -f1 -depsc -r300 WSP_vs_AWS_contour.eps']);
```