Experimental evolution reveals hidden diversity in evolutionary pathways

Peter A Lind, Andrew D Farr, Paul B Rainey

DOI: http://dx.doi.org/10.7554/eLife.07074

Cite as: eLife 2015;10.7554/eLife.07074

Received: 17 February 2015
Accepted: 24 March 2015
Published: 25 March 2015

22-03-2015

Experimental evolution reveals hidden diversity in evolutionary pathways

Peter A. Lind[a,*,§], Andrew D. Farr[a] and Paul B. Rainey[a,b]

[a] New Zealand Institute for Advanced Study and Allan Wilson Centre for Molecular Ecology and Evolution, Massey University at Albany, Auckland, 0745, New Zealand

[b] Max Planck Institute for Evolutionary Biology, Plön 24306, Germany

§ Current address: Dept. Medical Biochemistry and Microbiology, Uppsala University, Box 582, SE-75123, Sweden.

* Correspondence and requests for materials should be addressed to Peter A. Lind Dept. Medical Biochemistry and Microbiology, Uppsala University, Box 582, SE-75123, Sweden. E-mail: peter.lind@imbim.uu.se Phone: +46 (18) 4714604

Keywords: bacterial evolution, parallel evolution, genetic constraint, diguanylate cyclase, c-di-GMP, evolutionary rules

Author contributions: PAL Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article. ADF Conception and design, Acquisition of data, Analysis and interpretation of data. PBR Conception and design, Analysis and interpretation of data, Drafting or revising the article.

Abstract

Replicate populations of natural and experimental organisms often show evidence of parallel genetic evolution, but the causes are unclear. The wrinkly spreader morph of *Pseudomonas fluorescens* arises repeatedly during experimental evolution. The mutational causes reside exclusively within three pathways. By eliminating these, 13 new mutational pathways were discovered with the newly arising WS types having fitnesses similar to those arising from the commonly passaged routes. Our findings show that parallel genetic evolution is strongly biased by constraints and we reveal the genetic bases. From such knowledge, and in instances where new phenotypes arise via gene activation, we suggest a set of principles: evolution proceeds firstly via pathways subject to negative regulation, then via promoter mutations and gene fusions, and finally via activation by intragenic gain-of-function mutations. These principles inform evolutionary forecasting and have relevance to interpreting the diverse array of mutations associated with clinically identical instances of disease in humans.

Impact Statement

Genetic architecture governs the evolvability of adaptive paths providing a framework for evolutionary forecasting.

Introduction

Prediction of evolutionary change from a set of first principles, even in the most elementary of biological systems, has proven difficult (de Visser and Krug, 2014). This is due in part to the stochastic nature of mutation, but also to lack of understanding of the molecular properties of gene products and their interactions, i.e., the processes underpinning development of phenotypes – including genetic and developmental constraints – which are themselves a product of the genotype-to-phenotype map (Pigliucci, 2010).

The ubiquity of parallel genetic evolution, observed both in nature and in laboratory experiments (Gerstein et al., 2012, Meyer et al., 2012, Zhen et al., 2012, Flowers et al., 2009, Herron and Doebeli, 2013, Stern, 2013) shows that evolution can be highly reproducible. While repeated evolution of similar traits is considered evidence of adaptive evolution, it also suggests the possibility that evolution may be governed by rules that if understood, would lead to a more predictive science (Neher et al., 2014, Hansen, 2006, Bull and Molineux, 2008, de Visser and Krug, 2014, Stern and Orgogozo, 2009).

Not all explanations for parallel evolution necessitate the existence of underlying rules. If evolution proceeds via a single route – because there is no other – then there is no reason to suppose that evolution is anything other than idiosyncratic (Vogwill et al., 2014, Zhen et al., 2012, Jost et al., 2008). Should evolution proceed along a single pathway when multiple are available and yet the fitness of the phenotype from the common path is superior, then – other than the pleasure of discovery – there is no dilemma to solve. If, however, evolution proceeds along a single pathway and yet that pathway is just one of a number of possible routes to a range of phenotypes with equivalent fitness, then determining the underlying causes

becomes a matter of interest. Unfortunately opportunities for discovery of such pathways are limited (Gompel and Prud'homme, 2009, Stern, 2013) (Figure 1A).

One way to proceed is via model populations amenable to replication. The outcome of evolution in each population – propagated under identical conditions – can then be determined. Discovery of a novel solution would indicate the existence of at least two pathways – moreover, the relative fitness of types can be determined. While in principle straightforward, progress requires the analysis of many thousands of replicate populations (Desai, 2013) – or molecules (Ellington, 1994) – and even if possible, failure to find an alternate route would not mean that one does not exist.

An alternative approach is to take an experimental system where the most common genetic (mutational) pathways to a particular adaptive phenotype can be identified and then eliminated without deleterious effects on fitness of the ancestral type (Heineman et al., 2009). Thus evolution can be re-run from different starting genotypes that differ in the spectrum of pathways available to evolution (Figure 1B). If the same phenotypic solution can arise from a genotype devoid of the typically used genetic route, then it is clear that evolution has more options at its disposal that are typically realised.

Precisely this approach was taken previously using experimental populations of *Pseudomonas*. McDonald et al. revealed (McDonald et al., 2009) the existence of three commonly used mutational routes to a single adaptive "wrinkly spreader" (WS) phenotype, but showed that additional, less frequently utilised pathways existed (the nature of these pathways was not determined). Here, beginning with the ancestral type devoid of the three known routes to WS, we propagated multiple independent populations and identified 91 new WS mutants with similar fitness to the common WS types. A combination of genetics and genome sequencing revealed ten new single

mutational step routes and three additional paths requiring two or more mutations. Our data provide an explanation for why the newly discovered pathways are rarely followed, provide a set of hierarchical principles, and show how genetic constraints can bias the outcome of evolution.

Results

Experimental rationale

The *Pseudomonas fluorescens* SBW25 (Silby et al., 2009), hereafter "SBW25", experimental system of adaptive radiation has been extensively studied (Spiers et al., 2002, Bantinaki et al., 2007, McDonald et al., 2009, Rainey and Travisano, 1998) and has several features that makes it ideal for addressing the question of bias in evolutionary pathways. Every time the ancestral SM genotype of SBW25 is placed in a nutrient-rich static microcosm, metabolism-driven depletion of oxygen imposes strong selection for mutants that colonize the oxygen replete air-liquid interface (Figure 2A). The most successful of various mat-forming types (Ferguson et al., 2013) display a wrinkled morphology on agar plates (Figure 2A) and are known collectively as wrinkly spreaders (WS). In a previous experiment the mutational origins of 26 independent WS genotypes were unraveled (McDonald et al., 2009). All mutations resided in one of three pathways (Wsp, Aws and Mws). Each pathway harbors a di-guanylate cyclase (DGC) responsible for production of cyclic-di-GMP (c-di-GMP). When the cognate DGC is constitutively activated cells over-produce an acetylated cellulose polymer (Spiers et al., 2002, Spiers et al., 2003) – the proximate cause of the WS phenotype (Figure 2B) (McDonald et al., 2009). That evolution followed just three pathways was unexpected given that the SBW25 genome carries 39 putative DGCs (McDonald et al., 2009). The majority of mutations were loss-of-

function changes in negative regulators of the *wsp*, *aws* and *mws*-encoded DGCs (McDonald et al., 2009). Such a spectrum of mutations made sense given that loss-of-function mutations in each negative regulator resulted in constitutive activation of the DGC (Goymer et al., 2006, Malone et al., 2007), with ensuing downstream effects (Figure 2B) (McDonald et al., 2009). No evidence of mutational hotspots – a possible cause of bias – was obtained.

A unique feature of the experimental *Pseudomonas* model is ability to remove the common pathways from the ancestral SM type without deleterious effect on the fitness (McDonald et al., 2009). This allows a test of the hypothesis that there exist alternative evolutionary pathways to WS that are not typically followed, either because the WS types that are generated have low fitness, or because properties of the genotype-to-phenotype map make alternate routes unlikely.

Selection for new WS types and identifying mutations

200 independent glass microcosms were inoculated with the smooth (SM) Δ*wsp*Δ*aws*Δ*mws* mutant. After 6 d growth in static microcosms, dilutions were plated on KB agar and the resulting colonies screened for types exhibiting the WS morphology. WS types were found in 91 microcosms, in contrast, when the founding genotype is ancestral SM SBW25, all microcosms harbor WS types after just three days of propagation (McDonald et al., 2009).

The mutational causes of WS types were sought by suppressor analysis, using a transposon mutagenesis screen to find candidate loci for targeted Sanger sequencing (Giddens et al., 2007), or by genome re-sequencing. We found single mutations in 86 of the mutants in ten different loci, all encoding a protein with a putative DGC domain, similar to the three previously known pathways. Four of the remaining WS

types had double mutations and one genotype had three mutations. The mutational targets are summarized in Figure 3 and Table 1 and full details are available in Figure 3 – source data 1.

Intragenic mutations

The most common of the previously unseen routes to WS (43/91) involved mutation in PFLU0085, a putative DGC lacking other annotated domains. All 22 (7 unique) base pair substitutions (BPSs) were clustered in the region 1223-1340 of the open reading frame upstream of the DGC domain. Further disruptive mutations within PFLU0085 were identified including in-frame deletions (20 mutants, 3-477 bp) and a 141-bp duplication – all in the same 1223-1340 region – suggesting that a wide variety of disruptive mutations within a small window can produce WS. This implicates 1223-1340 as a negative regulator of the downstream DGC domain.

Four other genes harbored intragenic mutations, but these were less common, suggesting a smaller target size less consistent with a negative regulatory role. One mutation (R321L) was found in PFLU0956 close to the third predicted transmembrane helix preceding the DGC domain. A similar pattern was evident at PFLU3448 where two mutations affecting the same amino acid (A200V and A200T) were identified. This residue resides close to the last of seven predicted transmembrane helices. At PFLU3571, two mutations were identified – both with W13R substitutions in the first of two transmembrane helices. A third mutation in PFLU3571 was found in the C-terminal end of the HAMP domain close to the N-terminal part of the DGC domain. The relative rarity of these mutations compared to those in PFLU0085 suggests a reduced target size that is again inconsistent with a negative regulatory role. More likely, these mutations bring about changes in protein-

protein interactions, changes in localization of diguanylate cyclases in the cell, or alterations in the relative orientation of domains. All such alterations could reasonably activate production of c-di-GMP (and thus cellulose) without increasing catalytic rate. The fourth gene – that encoding PFLU5960 – contained a mutation (D160G) in the DGC domain itself. The affected amino acid is close to the active site (amino acids 200-205) based on a Phyre2 structure prediction model (Kelley and Sternberg, 2009) of the protein (Supplementary file 1). Such a mutation likely increases catalytic activity of the imperfect GSDEF site.

Promoter mutations

Mutations were found in the upstream region of three DGC-encoding genes (Figure 3, Figure 3 –source data 1). The most common was associated with PFLU0956 where six BPSs and three indels in the -54 to -59 region relative to the start codon were identified. Six BPSs and one 14 bp duplication were found upstream of PFLU5698 and two indels were found upstream of PFLU1349. In all cases, the position of the mutation indicates no impact on the ribosomal binding site. The effect of a subset of these promoter mutations on transcription is described below.

Gene fusions

For three loci, deletion mutations caused fusions between an open reading frame encoding a DGC domain and an upstream gene. Such mutations could lead to an increase in transcription by promoter capture, or may change the functional or regulatory connections of the protein resulting in activation of the associated DGC. For PFLU0183 eight deletions were identified: each generating an in-frame fusion with a putative a fatty acid desaturase (PFLU0184) located upstream. Eight deletions

generating in-frame fusions were also found between PFLU4306 and PFLU4305. PFLU4306 encodes a DGC domain protein and the upstream gene encodes a putative L-lactate dehydrogenase. A third example was defined by a single mutation that fused the GGDEF domain protein PFLU4308 to PFLU4313 – the latter located upstream of the DGC and encoding a hypothetical protein. This last fusion involved deletion of four intervening genes. For two of the loci WS-generating mutations arose at frequencies similar to those in promoters, suggesting that gene fusions and promoter capture play a major evolutionary role in gene and or protein evolution.

Double and triple mutations

Remarkably, double mutations in the same two genes, PFLU0458 and PFLU4744 were found in three independent WS mutants. Both contain early frame shift mutations or stop codons suggesting these are loss-of-function mutations (Figure 3, Figure 3 –source data 1). PFLU0458 harbors a DGC domain but without a catalytic motif (suggesting loss of enzymatic activity) and an EAL domain typically involved in c-di-GMP degradation. Loss-of-function mutations suggest a role for PFLU0458 as a c-di-GMP degrading enzyme, which is also supported by data from *P. aeruginosa*, where the orthologue PA5017 (*dipA*, *pch*) lacks DGC activity, but is proficient for phosphodiesterase activity (Roy et al., 2012). PFLU4744 encodes the alginate biosynthesis transcriptional activator *algZ*/*amrZ* that has a previously known role in cellulose expression and biofilm formation (Giddens et al., 2007). The remaining double mutant harbored a mutation in PFLU0458, as above, and a second mutation upstream (-51) of the DGC-encoding protein PFLU0621.

The single triple mutant harbored mutations in PFLU4744 that were found in combination with PFLU0458 described above, but in this instance were combined

with mutations in two other regulatory proteins: PFLU4414 (*cheA*, chemotaxis histidine kinase) and PFLU4443 (*adnA*/*fleQ*, flagella activator, negative regulator of *algR*). A connection between PFLU4744/PFLU4443 and the cellulose biosynthetic (*wss*) operon is known (Giddens et al., 2007).

Reconstruction of mutations and phenotypic characterization

Representative intragenic mutations, promoter mutations and gene fusions (Figure 4) were reconstructed in the ancestral SM SBW25 background to confirm that they are the sole cause of the WS phenotype and also to rule out any dependency on the Δ*wsp*Δ*aws*Δ*mws* genetic background.

To confirm that the phenotypic basis of the selective advantage is similar to the previously described WS mutants (where activation of a DGC results in overproduction of cellulose (McDonald et al., 2009)) reconstructed mutants were stained with calcofluor and Congo red and their colony morphology examined. All single mutants were positive for calcofluor staining and colony morphologies with Congo red staining showed that all stained to a higher degree than the SM ancestor, but there were differences (Figure 4 – figure supplement 1). All reconstructed mutants colonized the air-liquid interface of static microcosms within 24 h (Figure 4 – figure supplement 2).

Mutations found in the double and triple mutants (Figure 5) were reconstructed individually and then combined. For the double and triple mutants, only combinations of the individual mutations with a PFLU4744 mutation were positive for calcofluor and Congo red binding (Figure 5 – figure supplement 1). All individual and combined reconstructed mutants colonized the air-liquid interface of static microcosms within

24 h with the exception of the genotype carrying solely the PFLU4443 E398* mutation (Figure 4 – figure supplement 2).

Fitness of single WS mutants

The preferential use of the Wsp, Aws and Mws pathways to generate WS types when starting from the ancestral SM genotype could be a consequence of the WS types arising via the newly discovered pathways having lower fitness compared to those arising from mutation in the three known pathways. To investigate if reduced fitness explained the data, two types of competitive fitness assays were performed using the reconstructed mutants.

The first assay focused on the capacity of the WS types to invade, from rare, a numerically dominant population of ancestral SM types marked with GFP. This mimics the situation where a random WS mutant occurs in a population and rises to high frequency without clonal interference from other WS mutants. The second assay involved a 1:1 competitive fitness assay in which each reconstructed WS type was directly pitted against one of the most fit and common of the WS mutants previously described – the so named LSWS type (WspF S301R) – also marked with GFP (McDonald et al., 2009). This assay focused on competitive performance at the air-liquid interface – a realistic situation given that several WS mutants are likely to be present in the large populations ($\approx 10^{10}$ cells) used in the evolution experiments. Thus the two assays focus on different aspects of the adaptive radiation that are not necessarily directly correlated, for example a mutant can be superior in initial attachment, but grow slowly at the air-liquid interface. In addition, fitness in static microcosms is frequency-dependent (Rainey and Travisano, 1998).

All WS types harboring single mutations invaded the SM population from rare with selection coefficients (s) of 0.28-0.59 per generation compared to the control SM type (s = 0) (Figure 4A). Three WS mutants caused by mutations in the commonly followed mutation pathways (*wspF*, *awsX* and *mwsR*) were included to see if lower invasion fitness was responsible for the rarity of the newly discovered mutational routes. All three produced high selection coefficients. Two newly discovered WS types containing mutations in PFLU5698 and PFLU5960 were significantly less fit than the LSWS ($P<0.01$, two tailed t-test). However, taken together, the low (in a few instances) invasion rate for WS generated by mutations in the newly discovered pathways cannot account for the rarity of the alternative WS types found here.

In competition with the high fitness common LSWS (*wspF* S301R) mutant (s = 0), seven of the 13 reconstructed WS types had significantly lower fitness ($P<0.01$, two tailed t-test), including all four of the PFLU0085 mutants (Figure 4B). One WS type harboring a mutation in PFLU1349 had significantly higher fitness (s = 0.039, $P<0.01$, two tailed t-test).

Fitness of double and triple mutants

In the invasion assay all mutants, except the PFLU0458 E45*, PFLU4443 E398*, and PFLU4443 PFLU4414, invaded the SM population successfully (Figure 5A) (two tailed t-test $p>0.01$ vs. the control). In competition assay, the fitness of WS types harboring two mutations were no different to WS types arising from a single mutation and were overall indistinguishable from the LSWS control strain (Figure 5B). The individual PFLU0458 I835S mutant has a significantly higher fitness in the competition assay (s = 0.039, two tailed t-test $P = 1.8 \times 10^{-4}$). The triple mutants and all combination of genotypes carrying constituent mutations had lower fitness than

LSWS in the competition assay, with the highest fitness recorded for the single PFLU4744 I44T mutant (Figure 5B). A similar finding came from the invasion assay where this mutant also showed high competitive ability. WS types with various combinations of single mutations ranked at the lower end of the fitness spectrum with the PFLU4443, PFLU4443/PFLU4414 and triple mutation WS types being no different from the SM control (two tailed t-test $P>0.01$, Figure 5A).

Transcriptional effects of promoter mutations and gene fusions

WS types carrying putative promoter mutations (PFLU0956, PFLU1349, PFLU5698 and PFLU0621) were expected to activate DGCs by increasing transcription. This is a possible mechanism of activation for the gene fusions (PFLU0183, PFLU4306 and PFLU4308) where new promoters may have been captured by the deletion event or may have resulted in transcriptional terminators being lost. To examine the effect on transcription of representative promoter and gene fusion mutants we performed quantitative PCR. The promoter mutations all increased transcription, up to 40-fold compared to SBW25 (Figure 6): the PFLU0621 mutant showed a more modest 5-fold increase. The latter result is consistent with the fact that this mutation alone does not generate the WS morphology (Figure 5 – figure supplement 1). WS types carrying gene fusion mutations (PFLU4305-4306 and PFLU4313-4308) also showed large increases in transcription with 23- and 219-fold increases respectively, supporting promoter capture as the mechanism of activation. The PFLU0184-0183 fusion showed a 3.9-fold increase in transcription, which is not significantly different ($P=0.09$, two tailed t-test) to a control PFLU0085 mutation (Figure 6).

Discussion

322 Understanding genetic evolution requires knowledge of the factors that affect the
323 translation of mutation into phenotypic variation. While much is known about the
324 nature of mutation (Drake et al., 1998), knowledge of how change in DNA sequence
325 is translated into phenotypic variation – the raw material for natural selection – is less
326 well understood (Gompel and Prud'homme, 2009). Of particular relevance is the
327 complex network of functional and regulatory connectivities that define the genotype-
328 to-phenotype map. This network of interactions constrains (channels) evolution;
329 restricts the pathways it takes and – by imposing limits to phenotype space – defines
330 the rules by which it works (Hansen, 2006, Stern and Orgogozo, 2009, Gompel and
331 Prud'homme, 2009). The notion that evolution follows a limited subset of pathways
332 and conforms to rules is not new (Vavilov, 1922, Geoffroy Saint-Hilaire, 1818, Smith
333 et al., 1985), however, its relevance has often been challenged, largely through too
334 frequent invocation of constraints to explain negative data and through lack of
335 experimental insight (Brakefield and Roskam, 2006).
336 Our unique experimental system has allowed unveiling of hitherto hidden
337 evolutionary pathways by which the WS phenotype can be achieved. For the most
338 part WS types arising via the new pathways do not differ in fitness relative to known
339 WS types and therefore their rarity cannot be attributed to a selective disadvantage.
340 Genetic constraints, however, provide plausible explanations – certain pathways have
341 a greater capacity than others to translate mutation into phenotypic variation (Figure
342 7). While bias is well known to arise as a consequence of localized mutational
343 hotspots, the spectrum of mutations underpinning different routes to WS suggests a
344 minor influence of any such bias (Figure 3 Source data 1 (McDonald et al., 2011)).
345 Conceivably, the presence of a mutational hot spot at the exact location of an
346 intragenic activating mutational site or promoter could have a major impact on the

rate of phenotypic production from that locus, but the diversity of mutants obtained for these mutational types suggest that such biases are not dominant in this system.

All newly revealed mutational routes to the WS phenotype harbor proteins predicted to encode DGCs, all show the distinctive WS colony morphology and niche preference, and all over-produce cellulose. Despite underlying molecular similarity only three are routinely traveled: Wsp, Aws and Mws. The newly discovered mutational spectra allows proposals regarding probable mechanisms for activation of the pathways specific to each DGC and with this a more complete understanding of why genetic evolution prefers certain pathways over others.

From a purely genetic perspective there are no surprises: loss-of-function mutations are vastly more common that gain-of-function mutations (Gompel and Prud'homme, 2009, Lee et al., 2012, Herron and Doebeli, 2013) and thus those pathways containing DGCs subject to negative regulation will, by virtue of target size (a product of length of DNA and function), translate mutation into WS variation more efficiently than those pathways containing DGCs subject to other forms of regulation. But our detailed analysis makes possible a finer scale of resolution.

In common with the previously known Wsp, Aws and Mws pathways, PFLU0085 also appears – on the basis of the spectrum of DGC-activating mutations – to be subject to negative regulation. That WS arising via the PLU0085 are not detected when the founding genotype is ancestral SM is readily understood based on the small mutational target (~117 nucleotides) and the fact that mutations within this region cannot disrupt the reading frame else DGC activity is lost.

The second most common set of routes to WS, with 5-40 fold fewer mutations per gene compared to PFLU0085, are those involving promoter mutations and gene fusions (Figure 3). It is not clear why promoter mutations were only found for a

minority of the 39 DGC domain-containing proteins. Possibly this reflects the need for mutations to generate a very high level of transcription, but it is also consistent with the fact that regulation of many DGC-containing proteins is post-translational (Goymer et al., 2006, Jenal and Malone, 2006).

The high frequency of gene fusions is surprising given the relatively few reports of this kind of mutational event in experimental evolution studies, with the notable exception of promoter capture that was central to the evolution of an *E. coli* mutant that gained the ability to grow on citrate (Blount et al., 2012). The loss of genetic material through beneficial gene fusions could contribute to deletional bias over evolutionary timescales as observed in bacteria (Mira et al., 2001) and presents a selectionist alternative to reductive evolution by genetic drift or loss of biosynthetically expensive genes (Lee and Marx, 2012, Moran, 2002). Moreover, beneficial promoter capture events can create new protein domain combinations and provide raw material for the evolution of genes with novel functions, even if initially selected only because of a difference in transcriptional regulation.

The third most common mutation route to WS involves rare activating mutations that differ from intragenic negative regulators in that only specific base pair substitutions can lead to activation (Figure 3). Effects at the molecular level are unknown, but it appears that mutations modifying the active site are rare, given that only one mutation was found in proximity to the catalytic site. Assuming that any of the newly discovered pathways are used at a frequency that is at least ten-fold lower than for the common pathways to WS, it is possible to estimate that only one in one thousand activating mutations would directly modify the active site (Figure 7). This strongly supports the theory that the majority of advantageous mutations are likely to occur in regulatory regions, including promoters and regulatory proteins, or in the

case of intragenic mutations, primarily in regions peripheral to the main catalytic domain (Wray, 2007, McAdams et al., 2004).

WS types arising as a consequence of two mutations arose at a similar frequency as the rare activating mutations. Such double (and triple) mutants are unexpected given the seeming improbability of such events based on estimates of the genome mutation rate (Drake et al., 1998). Nonetheless their detection is not uncommon (Drake, 2007) and they have been explained by the existence of transient phenotypic mutators in the population due to transcriptional or translational errors in the expression of key genes involved in replication or DNA repair (Drake, 2007, Ninio, 1991). This suggests that once the third most common pathways involving single mutations have been realized, the same evolutionary principles can be applied to double mutations starting with double mutations in negative regulators, followed by double mutations in one negative regulator and one promoter mutation (Figure 3, Table 1).

Taken together our findings provide the clearest evidence yet that the network of regulatory interactions and connectivities that define the genotype-to-phenotype map directly affects the translation of mutation into phenotypic variation and that this can profoundly bias the course of genetic evolution. This has a number of implications. Firstly, our findings contribute to the growing number of studies that show that parallel phenotypic evolution is often underpinned by parallel genetic changes (see (Stern, 2013) Table 1 for a summary) (Zhen et al., 2012, Gerstein et al., 2012, Herron and Doebeli, 2013, Jost et al., 2008, Meyer et al., 2012, Blount et al., 2012)). However, whereas it is common to attribute such genetic parallelism to selection, our work suggests the need for caution: genetic architecture can be a significant contributory factor. It is even conceivable, within the bounds of

population parameters, such as population size and mutation rate, that certain high fitness phenotypes are never realized because of evolutionary bias.

Secondly, the WS-based experimental system has allowed discovery of pathways that evolution would rarely ever follow and suggests a hierarchical set of rules. These rules are consistent with the concept of 'target size' (Hansen, 2006, Gompel and Prud'homme, 2009), which can be formulated more specifically in terms of size of the gene and the likelihood that changes generate viable phenotypes. The latter depends on the opportunity for loss-of-function mutations to generate adaptive phenotypes, which, in turn, depends on the function of the gene. But it is clear that while evolution will proceed most readily via loss-of-function mutations (where possible), other kinds of genes that afford a much reduced target size cannot be overlooked. As we demonstrate, there exists opportunity for promoter mutations, gene fusions and activating mutations to also contribute to new phenotypes, but with a greatly reduced likelihood.

The existence of multiple genetic routes to a particular phenotypic end point is, given growing evidence of redundancy and evolvability in regulatory systems, possibly more common than currently appreciated (Stern, 2013, Gompel and Prud'homme, 2009, Heineman et al., 2009). If so, then the kinds of mutational patterns unraveled here may be evident elsewhere. We cautiously suggest that in addition to relevance for interpreting patterns of mutation underlying numerous studies of parallel and convergent evolution in natural and laboratory populations (Stern, 2013, Zhen et al., 2012, Gerstein et al., 2012, Herron and Doebeli, 2013, Jost et al., 2008, Meyer et al., 2012, Blount et al., 2012), our findings may also have value in understanding the spectrum of rare and common genetic variants underlying specific human diseases (Gibson, 2011). In the context of cancer, the mutational

heterogeneity of different cancer types is well documented and it is possible that the common (and less common) mutational trajectories (Lawrence et al., 2014, Yates and Campbell, 2012) make sense in light of constraints due to genetic architecture.

We have made little of our findings in terms of the regulatory networks underpinning c-di-GMP synthesis and cellulose expression in bacteria – including mechanisms of DGC activation – but the degree of redundancy (under one set of laboratory conditions) and capacity for evolutionary change by recruitment of different DGC-containing proteins and pathways is remarkable. The evident flexibility suggests that evolutionary rearrangement of interacting modules likely occurs in natural populations – it also underpins recent experimental findings on the evolution of multicellular life cycles (Hammerschmidt et al., 2014). Indeed, such facility for re-wiring may partly explain the diverse modular arrangement of both DGC domain-encoding genes and DGC-containing operons found among even closely related strains of a single species (Jenal and Malone, 2006).

The hierarchical rules for the evolution of new WS phenotypes revealed through the analysis of experimental *Pseudomonas* populations may be sufficiently general for them to be applied to other biological settings where an adaptive challenge can be solved by mutational gene activation. Such settings include the evolution of antibiotic resistance, the evolution of virulence in pathogens and the emergence of complex traits in eukaryotic populations. However, the relevance of these principles remains to be tested. This could occur via *a priori* predictions of mutational routes for specific populations in given selective contexts. If robust, then the claim to have moved closer to a more predictive theory of genetic evolution will have substance.

Materials and Methods

Strains and media

All strains used are *Pseudomonas fluorescens* SBW25 (Silby et al., 2009) or derivatives thereof except for *E. coli* strains used for strain construction and transposon mutagenesis (*E. coli* DH5-α λ_{pir}, *E. coli* SM10 λ_{pir} IS-Ω-kan/hah, *E.coli* pRK2013). *P. fluorescens* strains were grown in King's medium B (KB) (King et al., 1954) at 28°C and *E. coli* strains were grown in lysogeny broth (LB) (Bertani, 1951) at 37°C. Solid media was KB or LB with 1.5% agar.

Antibiotics concentrations for strain construction and plasmid maintenance were gentamycin (10 mg/l), kanamycin (100 mg/l), tetracycline (15 mg/l), nitrofurantoin (100 mg/l) and cycloserine (1000 mg/l). X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) was used at a concentration of 40 mg/l in agar plates. Calcofluor (Fluorescent brightener 28) was added to agar plates at a concentration of 35 mg/l and Congo red at 10 mg/l.

Experimental evolution to select for WS

Individual colonies of the smooth ancestor strain PBR716 (Δwsp Δaws Δmws) (McDonald et al., 2009) was used to inoculate 200 glass microcosms (6 ml KB) which were incubated statically for 72 h at 28°C. The microcosms were then vortexed vigorously and diluted 1000 times into new KB microcosms that was incubated for another 72 h under identical condition before suitable dilutions were spread onto agar plates. After incubation for 48 h the plates were screened (500-2,000 colonies) for colonies with different morphology than the SM ancestor.

Transposon mutagenesis and sequencing of transposon insertion sites

Transposon mutagenesis was used to find candidate genes for alternative WS mutations as previously described (Giddens et al., 2007). Briefly, the plasmid pCM639 containing the IS-Ω-kan/hah transposon was conjugated from *E. coli* SM10 λ_{pir} into the recipient *P. fluorescens* WS strain using an *E. coli* pRK2013 helper strain. Suitable dilutions of successful transconjugants were selected on KB plates with kamamycin for selection of the transposon and nitrofurantoin for counterselection of *E. coli*. Fewer than 1000 transconjugants from each independent conjugation were screened for loss of the WS colony morphology and single colonies were isolated on agar plates. The insertion sites in the genome were found by an arbitrarily primed PCR approach and Sanger sequencing (Macrogen) of the products (Manoil, 2000).

DNA sequencing to find alternative WS mutations

Candidate genes from the transposon suppressor analysis were sequenced by Sanger sequencing (Macrogen) in all alternative WS in an iterative fashion, eliminating the common pathways before moving on to the next round of transposon mutagenesis and sequencing. For a few mutants that consistently failed to produce suppressors, were difficult to phenotypically distinguish from the ancestor or had low conjugation efficiency we used genome sequencing to find the mutations. Genomic DNA was prepared using the Wizard® Genomic DNA purification kit (Promega), sequenced by the Australian Genome Research Facility using Illumina HiSeq2000 and assembled against the reference *P. fluorescens* SBW25 genome using Geneious 5.5.6 (Biomatters). All oligonucleotide primers used in this study are available in Supplementary file 2.

Reconstruction of mutations in SBW25

Representative intragenic mutations (PFLU0085 L408P, V447G, ΔR361-R414, ΔR437-A458; PFLU3448 A200T; PFLU3571 W13R; PFLU5960 D160G), promoter mutations (PFLU0956 T-54G; PFLU1349 ins TC -47/48, PFLU5698 C-73T) and gene fusions (PFLU0184 M1-T328 fused to PFLU0183 A29-G335; PFLU4305 M1-Y340 fused to PFLU4306 S21-G489; PFLU4313 M1-F115 fused to PFLU4308 A189-R820) were reconstructed in the SBW25 background to prove that they are the cause of the WS phenotype and to demonstrate that these mutational pathways are available in the ancestral strain and not dependent on deletion of the *wsp*, *aws* and *mws* loci. Mutations found in the double and triple mutants (PFLU0458 I835S, PFLU0621 -51 C>T; PFLU0458 E45*, PFLU4744 S39N; PFLU4414 Y652fs, PFLU4443 E398*, PFLU4744 I44T) were reconstructed individually and then combined. We used a two-step allelic replacement method to transfer the mutations into the ancestral background as previously described (Rainey, 1999, Bantinaki et al., 2007). In summary, PCR (Phusion High-Fidelity DNA polymerase, Thermo Scientific) was used to amplify an approximately 1000 bp region surrounding each mutation and the product was subsequently cloned into the pCR8 plasmid and sequenced. The cloned fragment from pCR8 was then moved to the pUIC3 suicide vector and mobilized into *P. fluorescens* SBW25 were it integrates into the chromosome by homologous recombination. After non-selective growth in KB, 10 mg/l tetracycline was added to inhibit the growth of cells that had lost the pUIC3 insert. After 2h 1000 mg/ml cycloserine was added to kill growing cells and enrich for cells that had lost pUIC3 with the tetracycline resistance marker and incubated for 4h at 28°C. Suitable dilutions were then plated on agar plates with X-gal to allow screening for loss of the *lacZ* gene on pUIC3. White colonies were confirmed to be tetracycline sensitive and single colonies were isolated. The region containing the

desired mutation was then sequenced in a number of colonies, with both the ancestral and wrinkly phenotype to exclude picking bias and confirm that the wrinkly phenotype was linked to the mutation.

Construction of competition strains

We used the wild type SBW25 and a previously described high fitness WS mutant LSWS (*wspF* A901C, S301R) (Bantinaki et al., 2007, McDonald et al., 2009) to construct Green Fluorescent Protein (GFP) expressing strains for use in competition experiments to determine the relative fitness of the alternative WS. These strains were genetically tagged in the chromosome with a mini-Tn7 transposon expressing GFP and a gentamicin resistance marker (miniTn7(Gm)P_{rrnB} P1 gfp-a) (Lambertsen et al., 2004) that was transferred from *E. coli* by conjugation together with the pUX-BF13 plasmid carrying the transposase genes.

Competition fitness assay

Competitive fitness was determined relative to the LSWS strain (Spiers et al., 2002, Goymer et al., 2006) (SBW25 *wspF* A901C, S301R) marked with GFP. Strains were grown for 16 h, shaking, in KB at 28°C before they were mixed at equal volumes diluted six times and grown for 4 h at the same conditions to ensure that the strains were in the same physiological state before the competition started. The initial ratios of alternative WS to LSWS GFP were determined by counting 100,000 cells using flow cytometry (BD FACS Canto) detecting GFP fluorescence on the 488 nm laser with 530/30 bandwidth filter. Suitable dilutions of the initial population were plated on KB agar plates to determine viable counts. The mix of alternative WS and LSWS was diluted 1000-fold in KB and incubated for 24 h, static at 28°C. Final viable

571 counts and ratios was determined in the same way. After incubation for approximately
572 40 h at 28°C viable counts was determined and the stability of the WS phenotypes and
573 GFP marker was determined. This was possible as most WS types have slightly
574 different colony morphologies, which allow us to determine if the GFP marker was
575 lost and the emergence of smooth ancestral types is easily detected. Rarely, smooth
576 colonies were found and if they made up more than 5% of the total population the
577 competition data was discarded for that microcosm. The number of generations was
578 determined by ln(final population/initial population)/ln(2). Selection coefficients were
579 calculated using the regression model s = [ln(R(t)/R(0))]/[t], as previously described
580 (Dykhuizen, 1990) where R is the ratio of alternative WS mutant to LSWS GFP and t
581 the number of generations. Control experiments with LSWS vs. LSWS GFP were
582 performed to compensate for the fitness cost (s = 0.06 +/- 0.01) of the miniTn7 with
583 the GFP marker. For each strain the competition assay was performed in
584 quadruplicates at a minimum of two separate occasions.
585
586 **Invasion fitness assay**
587 Invasion fitness was measured relative to the smooth (SM) ancestral SBW25 mini-
588 Tn7 GFP. Invasion strains were grown in KB for 24 h shaking and then mixed 1:100
589 with the SM GFP strain and a 1000-fold dilution of this mix was used to inoculate 6
590 ml static microcosms. After 48 h at 28°C the ratio of unmarked WS to GFP marked
591 SM was determined by flow cytometry as described above. Viable counts on KB
592 plates of initial and final populations were performed to calculate the number of
593 generations during the invasion growth. The stabilities of the GFP marker and colony
594 morphologies were confirmed and data from microcosms with >5% wrinkly GFP or
595 smooth unmarked were discarded. Selection coefficients relative to the ancestral

SBW25 strain were calculated as described above with compensation for the cost of the GFP marker determined by control invasion experiments with SBW25 vs. SBW25 GFP. The averages presented are the result of at least two independent experiments with quadruplicates for each strain.

Quantitative RT-PCR

Total RNA was isolated from reconstructed WS mutants with mutations in upstream promoter regions or fusions to other open reading frames using the SV Total RNA Isolation system (Promega). Cells were harvested at OD_{600} = 0.6 and resuspended in Tris-1mM EDTA buffer with 0.4 mg/ml lysozyme to lyse the cells before using the supplied protocol. The RNA was reverse transcribed into cDNA using the High Capacity cDNA reverse transcription kit (Applied Biosystems) and diluted 40 times before use in quantitative real-time PCR (DyNAmo Colorflash SYBR Green qPCR kit (Thermo Scientific), PikoReal 96 Real-Time PCR System (Thermo Scientific)). Relative changes in mRNA levels between the ancestral SBW25 strain and the WS mutants were determined with the ΔΔCq method using *recA* as an internal control (Livak and Schmittgen, 2001). We used two biological replicates per strain and three technical replicates per run over at least two separate experiments.

Acknowledgements

Supported by the Marsden Fund Council from New Zealand Government funding, administered by the Royal Society of New Zealand.

References

BANTINAKI, E., KASSEN, R., KNIGHT, C. G., ROBINSON, Z., SPIERS, A. J. & RAINEY, P. B. 2007. Adaptive divergence in experimental populations of Pseudomonas fluorescens. III. Mutational origins of wrinkly spreader diversity. *Genetics,* 176**,** 441-53.

BERTANI, G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *Journal of bacteriology,* 62**,** 293-300.

BLOUNT, Z. D., BARRICK, J. E., DAVIDSON, C. J. & LENSKI, R. E. 2012. Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature,* 489**,** 513-8.

BRAKEFIELD, P. M. & ROSKAM, J. C. 2006. Exploring evolutionary constraints is a task for an integrative evolutionary biology. *Am Nat,* 168 Suppl 6**,** S4-13.

BULL, J. J. & MOLINEUX, I. J. 2008. Predicting evolution from genomics: experimental evolution of bacteriophage T7. *Heredity (Edinb),* 100**,** 453-63.

DE VISSER, J. A. & KRUG, J. 2014. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet,* 15**,** 480-90.

DESAI, M. M. 2013. Statistical questions in experimental evolution. *Journal of Statistical Mechanics-Theory and Experiment*.

DRAKE, J. W. 2007. Too many mutants with multiple mutations. *Crit Rev Biochem Mol Biol,* 42**,** 247-58.

DRAKE, J. W., CHARLESWORTH, B., CHARLESWORTH, D. & CROW, J. F. 1998. Rates of spontaneous mutation. *Genetics,* 148**,** 1667-86.

DYKHUIZEN, D. E. 1990. Experimental studies of natural selection in bacteria. *Annu. Rev. Ecol. Syst.,* 21**,** 373-398.

ELLINGTON, A. D. 1994. Empirical Explorations of Sequence Space - Host-Guest Chemistry in the Rna World. *Berichte Der Bunsen-Gesellschaft-Physical Chemistry Chemical Physics,* 98**,** 1115-1121.

FERGUSON, G. C., BERTELS, F. & RAINEY, P. B. 2013. Adaptive Divergence in Experimental Populations of Pseudomonas fluorescens. V. Insight into the Niche Specialist "Fuzzy Spreader" Compels Revision of the Model Pseudomonas Radiation. *Genetics*.

FLOWERS, J. M., HANZAWA, Y., HALL, M. C., MOORE, R. C. & PURUGGANAN, M. D. 2009. Population genomics of the Arabidopsis thaliana flowering time gene network. *Molecular biology and evolution,* 26**,** 2475-86.

GEOFFROY SAINT-HILAIRE, E. 1818. *Philosophie anatomique,* Paris,, J.B. Baillière.

GERSTEIN, A. C., LO, D. S. & OTTO, S. P. 2012. Parallel genetic changes and nonparallel gene-environment interactions characterize the evolution of drug resistance in yeast. *Genetics,* 192**,** 241-52.

GIBSON, G. 2011. Rare and common variants: twenty arguments. *Nat Rev Genet,* 13**,** 135-45.

GIDDENS, S. R., JACKSON, R. W., MOON, C. D., JACOBS, M. A., ZHANG, X. X., GEHRIG, S. M. & RAINEY, P. B. 2007. Mutational activation of niche-specific genes provides insight into regulatory networks and bacterial function in a complex environment. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 18247-52.

GOMPEL, N. & PRUD'HOMME, B. 2009. The causes of repeated genetic evolution. *Dev Biol,* 332**,** 36-47.

GOYMER, P., KAHN, S. G., MALONE, J. G., GEHRIG, S. M., SPIERS, A. J. & RAINEY, P. B. 2006. Adaptive divergence in experimental populations of Pseudomonas fluorescens. II. Role of the GGDEF regulator WspR in evolution and development of the wrinkly spreader phenotype. *Genetics,* 173**,** 515-26.

HAMMERSCHMIDT, K., ROSE, C. J., KERR, B. & RAINEY, P. B. 2014. Life cycles, fitness decoupling and the evolution of multicellularity. *Nature,* 515**,** 75-9.

HANSEN, T. F. 2006. The evolution of genetic architecture. *Annual Review of Ecology Evolution and Systematics,* 37**,** 123-157.

HEINEMAN, R. H., BULL, J. J. & MOLINEUX, I. J. 2009. Layers of evolvability in a bacteriophage life history trait. *Mol Biol Evol,* 26**,** 1289-98.

HERRON, M. D. & DOEBELI, M. 2013. Parallel evolutionary dynamics of adaptive diversification in Escherichia coli. *PLoS biology,* 11**,** e1001490.

JENAL, U. & MALONE, J. 2006. Mechanisms of cyclic-di-GMP signaling in bacteria. *Annu Rev Genet,* 40**,** 385-407.

JOST, M. C., HILLIS, D. M., LU, Y., KYLE, J. W., FOZZARD, H. A. & ZAKON, H. H. 2008. Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. *Molecular biology and evolution,* 25**,** 1016-24.

688 KELLEY, L. A. & STERNBERG, M. J. 2009. Protein structure prediction on the
689 Web: a case study using the Phyre server. *Nature protocols,* 4**,** 363-71.
690 KING, E. O., WARD, M. K. & RANEY, D. E. 1954. Two simple media for the
691 demonstration of pyocyanin and fluorescin. *The Journal of laboratory and*
692 *clinical medicine,* 44**,** 301-7.
693 LAMBERTSEN, L., STERNBERG, C. & MOLIN, S. 2004. Mini-Tn7 transposons
694 for site-specific tagging of bacteria with fluorescent proteins. *Environmental*
695 *microbiology,* 6**,** 726-32.
696 LAWRENCE, M. S., STOJANOV, P., MERMEL, C. H., ROBINSON, J. T.,
697 GARRAWAY, L. A., GOLUB, T. R., MEYERSON, M., GABRIEL, S. B.,
698 LANDER, E. S. & GETZ, G. 2014. Discovery and saturation analysis of
699 cancer genes across 21 tumour types. *Nature,* 505**,** 495-501.
700 LEE, H., POPODI, E., TANG, H. & FOSTER, P. L. 2012. Rate and molecular
701 spectrum of spontaneous mutations in the bacterium Escherichia coli as
702 determined by whole-genome sequencing. *Proceedings of the National*
703 *Academy of Sciences of the United States of America,* 109**,** E2774-83.
704 LEE, M. C. & MARX, C. J. 2012. Repeated, selection-driven genome reduction of
705 accessory genes in experimental populations. *PLoS genetics,* 8**,** e1002651.
706 LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression
707 data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.
708 *Methods,* 25**,** 402-8.
709 MALONE, J. G., WILLIAMS, R., CHRISTEN, M., JENAL, U., SPIERS, A. J. &
710 RAINEY, P. B. 2007. The structure-function relationship of WspR, a
711 Pseudomonas fluorescens response regulator with a GGDEF output domain.
712 *Microbiology,* 153**,** 980-94.
713 MANOIL, C. 2000. Tagging exported proteins using Escherichia coli alkaline
714 phosphatase gene fusions. *Methods in enzymology,* 326**,** 35-47.
715 MCADAMS, H. H., SRINIVASAN, B. & ARKIN, A. P. 2004. The evolution of
716 genetic regulatory systems in bacteria. *Nature reviews. Genetics,* 5**,** 169-78.
717 MCDONALD, M. J., COOPER, T. F., BEAUMONT, H. J. E. & RAINEY, P. B.
718 2011. The distribution of fitness effects of new beneficial mutations in
719 Pseudomonas fluorescens. *Biology Letters,* 7**,** 98-100.
720 MCDONALD, M. J., GEHRIG, S. M., MEINTJES, P. L., ZHANG, X. X. &
721 RAINEY, P. B. 2009. Adaptive divergence in experimental populations of

Pseudomonas fluorescens. IV. Genetic constraints guide evolutionary trajectories in a parallel adaptive radiation. *Genetics,* 183**,** 1041-53.

MEYER, J. R., DOBIAS, D. T., WEITZ, J. S., BARRICK, J. E., QUICK, R. T. & LENSKI, R. E. 2012. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science,* 335**,** 428-32.

MIRA, A., OCHMAN, H. & MORAN, N. A. 2001. Deletional bias and the evolution of bacterial genomes. *Trends in genetics : TIG,* 17**,** 589-96.

MORAN, N. A. 2002. Microbial minimalism: genome reduction in bacterial pathogens. *Cell,* 108**,** 583-6.

NEHER, R. A., RUSSELL, C. A. & SHRAIMAN, B. I. 2014. Predicting evolution from the shape of genealogical trees. *Elife,* 3.

NINIO, J. 1991. Transient mutators: a semiquantitative analysis of the influence of translation and transcription errors on mutation rates. *Genetics,* 129**,** 957-62.

PIGLIUCCI, M. 2010. Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 365**,** 557-66.

RAINEY, P. B. 1999. Adaptation of Pseudomonas fluorescens to the plant rhizosphere. *Environmental microbiology,* 1**,** 243-57.

RAINEY, P. B. & TRAVISANO, M. 1998. Adaptive radiation in a heterogeneous environment. *Nature,* 394**,** 69-72.

ROY, A. B., PETROVA, O. E. & SAUER, K. 2012. The phosphodiesterase DipA (PA5017) is essential for Pseudomonas aeruginosa biofilm dispersion. *Journal of bacteriology,* 194**,** 2904-15.

SILBY, M. W., CERDENO-TARRAGA, A. M., VERNIKOS, G. S., GIDDENS, S. R., JACKSON, R. W., PRESTON, G. M., ZHANG, X. X., MOON, C. D., GEHRIG, S. M., GODFREY, S. A., KNIGHT, C. G., MALONE, J. G., ROBINSON, Z., SPIERS, A. J., HARRIS, S., CHALLIS, G. L., YAXLEY, A. M., HARRIS, D., SEEGER, K., MURPHY, L., RUTTER, S., SQUARES, R., QUAIL, M. A., SAUNDERS, E., MAVROMATIS, K., BRETTIN, T. S., BENTLEY, S. D., HOTHERSALL, J., STEPHENS, E., THOMAS, C. M., PARKHILL, J., LEVY, S. B., RAINEY, P. B. & THOMSON, N. R. 2009. Genomic and genetic analyses of diversity and plant interactions of Pseudomonas fluorescens. *Genome biology,* 10**,** R51.

SMITH, J. M., BURIAN, R., KAUFFMAN, S., ALBERCH, P., CAMPBELL, J., GOODWIN, B., LANDE, R., RAUP, D. & WOLPERT, L. 1985. Developmental Constraints and Evolution. *Quarterly Review of Biology,* 60**,** 265-287.

SPIERS, A. J., BOHANNON, J., GEHRIG, S. M. & RAINEY, P. B. 2003. Biofilm formation at the air-liquid interface by the Pseudomonas fluorescens SBW25 wrinkly spreader requires an acetylated form of cellulose. *Molecular microbiology,* 50**,** 15-27.

SPIERS, A. J., KAHN, S. G., BOHANNON, J., TRAVISANO, M. & RAINEY, P. B. 2002. Adaptive divergence in experimental populations of Pseudomonas fluorescens. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics,* 161**,** 33-46.

STERN, D. L. 2013. The genetic causes of convergent evolution. *Nature reviews. Genetics,* 14**,** 751-64.

STERN, D. L. & ORGOGOZO, V. 2009. Is genetic evolution predictable? *Science,* 323**,** 746-51.

VAVILOV, N. I. 1922. The law of homologous series in variation. *Journal of Genetics,* 12**,** 47-89.

VOGWILL, T., KOJADINOVIC, M., FURIO, V. & MACLEAN, R. C. 2014. Testing the role of genetic background in parallel evolution using the comparative experimental evolution of antibiotic resistance. *Mol Biol Evol,* 31**,** 3314-23.

WRAY, G. A. 2007. The evolutionary significance of cis-regulatory mutations. *Nature reviews. Genetics,* 8**,** 206-16.

YATES, L. R. & CAMPBELL, P. J. 2012. Evolution of the cancer genome. *Nat Rev Genet,* 13**,** 795-806.

ZHEN, Y., AARDEMA, M. L., MEDINA, E. M., SCHUMER, M. & ANDOLFATTO, P. 2012. Parallel molecular evolution in an herbivore community. *Science,* 337**,** 1634-7.

Figure 1. Determining the causes of parallel evolution. (A) A selective challenge repeatedly leads to a phenotypic adaptation with the same mutational solution represented by red circles. If only one solution is ever observed (the red circles), how can it be proven that alternative pathways exist (the unrealized (shaded) mutational

789 pathways marked with '?')? (B) In an experimental system where common genetic
790 pathways to an adaptive phenotype can be removed (the red circle with a cross),
791 alternative pathways (blue and yellow) can be revealed and the underlying causes of
792 the parallel evolution determined.
793
794 **Figure 2. The *Pseudomonas fluorescens* SBW25 wrinkly spreader model.** (A) The
795 ancestral smooth (SM) strain evolves to colonize the air-liquid interface of a static
796 microcosm. The ability to make a mat at the air-liquid interface is dependent on
797 activation of the gene products of the *wss* operon, which encodes the biosynthetic
798 machinery for production of cellulose that function as an extracellular glue (Spiers et
799 al., 2003). These mat-forming types display a wrinkled morphology on agar plates
800 and are referred to as wrinkly spreaders (WS). Over-activation of cellulose production
801 causes an increase in the second messenger c-di-GMP – the product of diguanylate
802 cyclases (DGCs) (Goymer et al., 2006, Malone et al., 2007). (B) All WS mutants
803 described in previous work have mutations in one of three loci (*wsp*, *aws* and *mws*),
804 all involving DGCs under negative regulation (McDonald et al., 2009)). There are 36
805 additional DGCs in the genome and the ability to remove the three commonly
806 followed pathways makes it possible to determine whether evolution can follow
807 alternate routes to the WS phenotype.
808
809 **Figure 3.** Mutational routes to WS. Numbers in parentheses are the number of
810 independent mutants found out of 91. Putative functional effects of mutations are
811 based on mutational patterns. Classification of mutational types has been colour
812 coded: Red, extragenic negative regulators; Orange, intragenic negative regulators;
813 Yellow, promoters (activating mutations); Green, gene fusion/promoter capture; Blue,
814 intragenic (activating mutations); and Purple, double and triple mutations.
815
816 **Figure 4. Fitness effects of WS types arising from single mutations.** (A) Fitness in
817 a 1:100 invasion assay against the ancestral SM genotypes. (B) Fitness in a 1:1
818 competitive assay against the high fitness LSWS (WspF S301R) genotype. Error bars
819 represent SD (n=8).
820
821 **Figure 5. Fitness effects of WS types arising from two and three mutational**
822 **events.** Individual mutations were recreated in the ancestral SM genotype plus

combinations of mutations (see text). (A) Fitness in a 1:100 invasion assay against the ancestral SM genotypes. (B) Fitness in a 1:1 competitive assay against the high fitness LSWS (WspF S301R) genotype. Error bars represent SD (n=8).

Figure 6. Relative increase of mRNA levels in WS mutants with putative promoter or gene fusion mutations compared to SM SBW25 measured by quantitative PCR. PFLU0085 is the Δ361-414 deletion mutant not expected to result in increased mRNA levels used as a control. Error bars represent SEM.

Figure 7. Hierarchical principles of mutational gene activation. Certain pathways have a greater capacity than others to translate mutation into phenotypic variation depending on the functional and regulatory interactions involved. The estimated number of mutants is a first approximation of how mutations are expected to be distributed between different categories assuming a similar number of target genes in each category. This distribution can also be biased by mutational hot spots and differences in fitness effects of mutations in the gene involved.

Table 1. Summary of mutations in different categories			
Category	**Gene locus**	**Types of mutation**	**Number of mutants**
Extragenic negative regulator	wspF (PFLU1224)	deletion, insertion, stop, frameshift, amino acid substitution	previous work*
	awsX (PFLU5211)	deletion, insertion, stop, frameshift, amino acid substitution	previous work*
Intragenic negative regulator	mwsR (PFLU5329)	deletion, insertion, amino acid substitution	previous work*
	PFLU0085	deletion, insertion, amino acid substitution	43
Promoter activating	PFLU0956	substitutions, small deletion and insertions	9
	PFLU5698	substitutions, small insertion	7
	PFLU1349	deletion, small insertion	2
Promoter capture / gene fusion	PFLU0183	deletion	8
	PFLU4306	deletion	8

	PFLU4308	deletion	1
Intragenic activating	PFLU0956	amino acid substitution	1
	PFLU3448	amino acid substitution	2
	PFLU3571	amino acid substitution	3
	PFLU5960	amino acid substitution	2
Double and triple mutants	PFLU0458	amino acid substitution, stop codon, frameshift	3
	PFLU4744	amino acid substitution, frame shift	
	PFLU0458	amino acid substitution	1
	PFLU0621	substitution (promoter)	
	PFLU0458	amino acid substitution	
	PFLU0621	substitution (promoter)	
	PFLU4414	Frame shift	1
	PFLU4443	stop	
	PFLU4744	amino acid subsitution	
* Mutational targets described in McDonald et al. (2009)			

Figure 4 – supplement 1. Colony morphology of WS types arising from single mutations on KB agar with and without Congo Red.

Figure 4 – supplement 2. Microcosms of WS types arising from single mutations after 24 h and 72 h static growth.

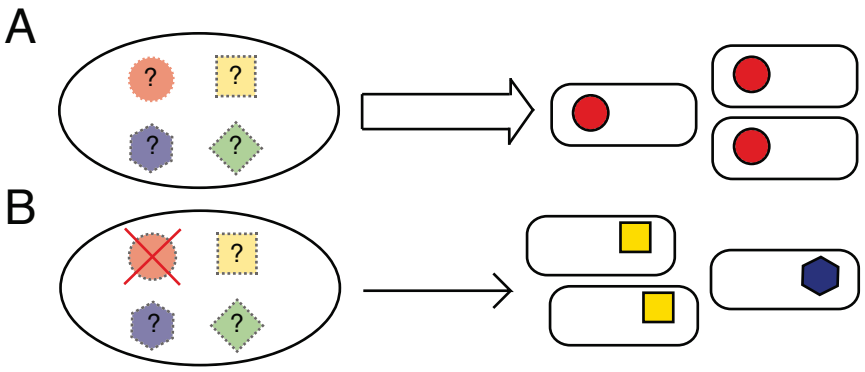
Figure 5 – supplement 1. Colony morphology of WS types arising from two and three mutational events on KB agar with and without Congo Red.

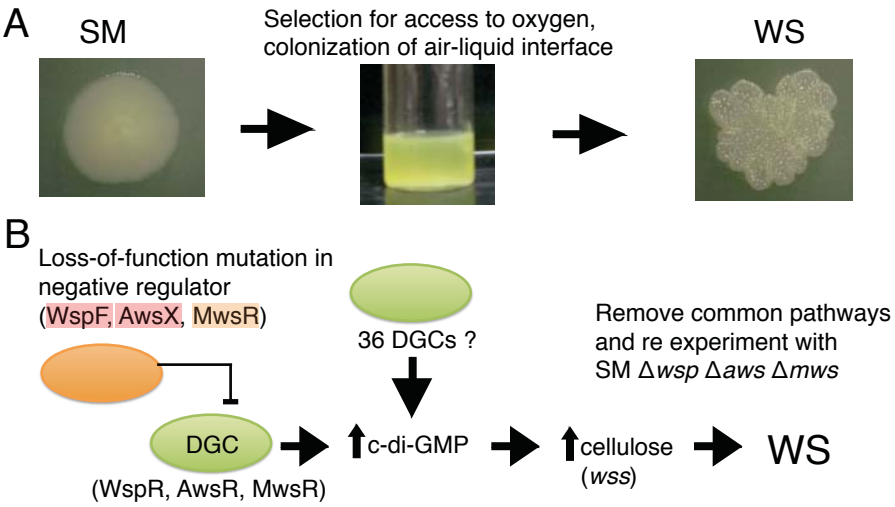
Figure 5 – supplement 2. Microcosms of WS types arising from two and three mutational events after 24 h and 72 h static growth.

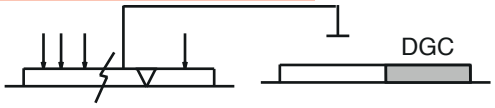
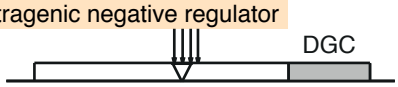
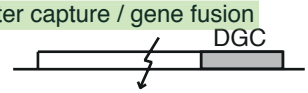
Figure 3-source data 1. Full genetic data for all new WS mutants.

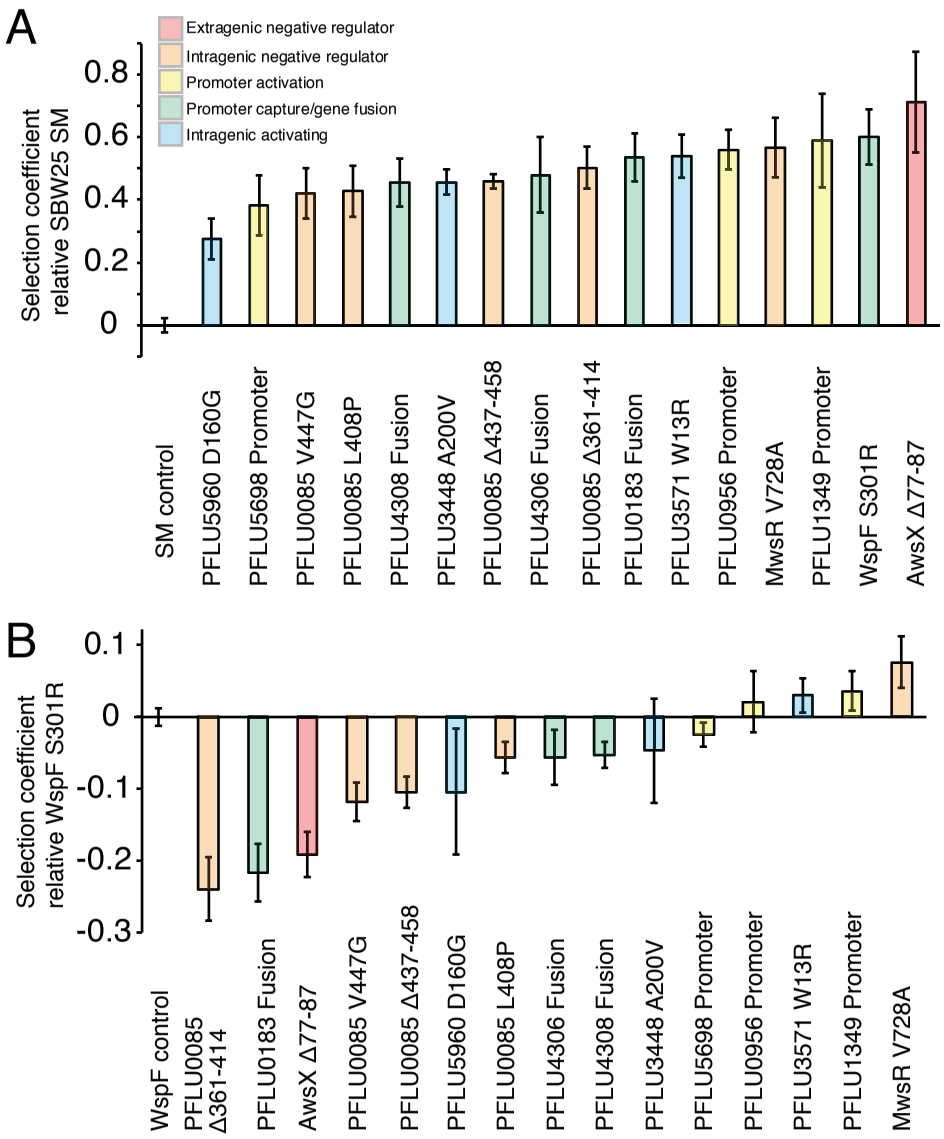
Supplementary File 1. Phyre2 pdb model of PFLU5960.

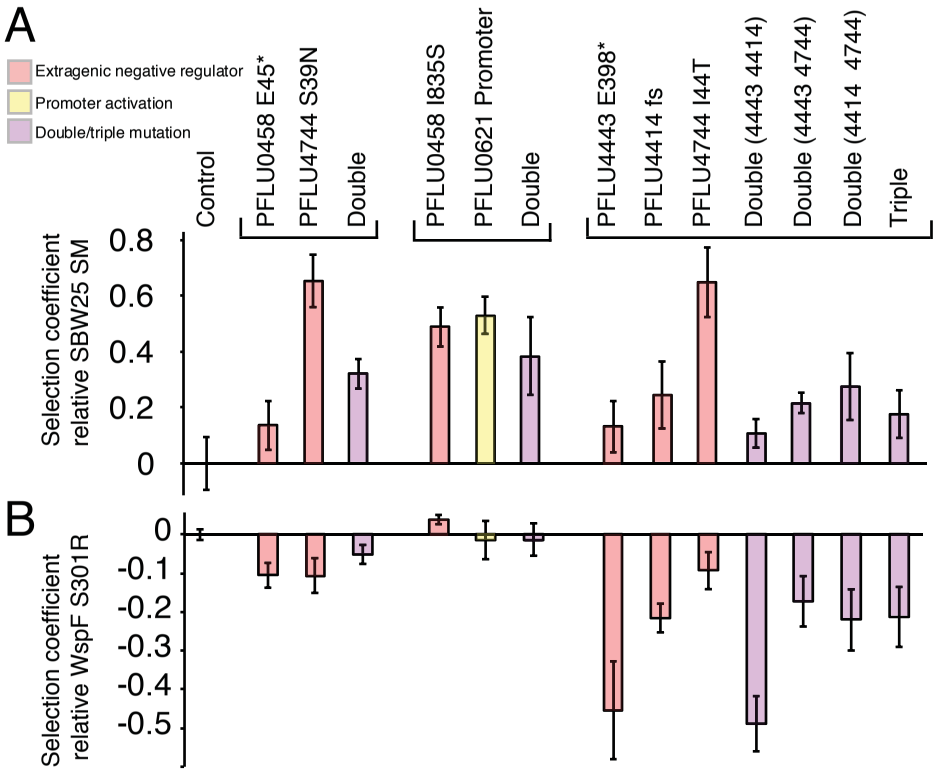
Supplementary File 2. Oligonucleotide primers.

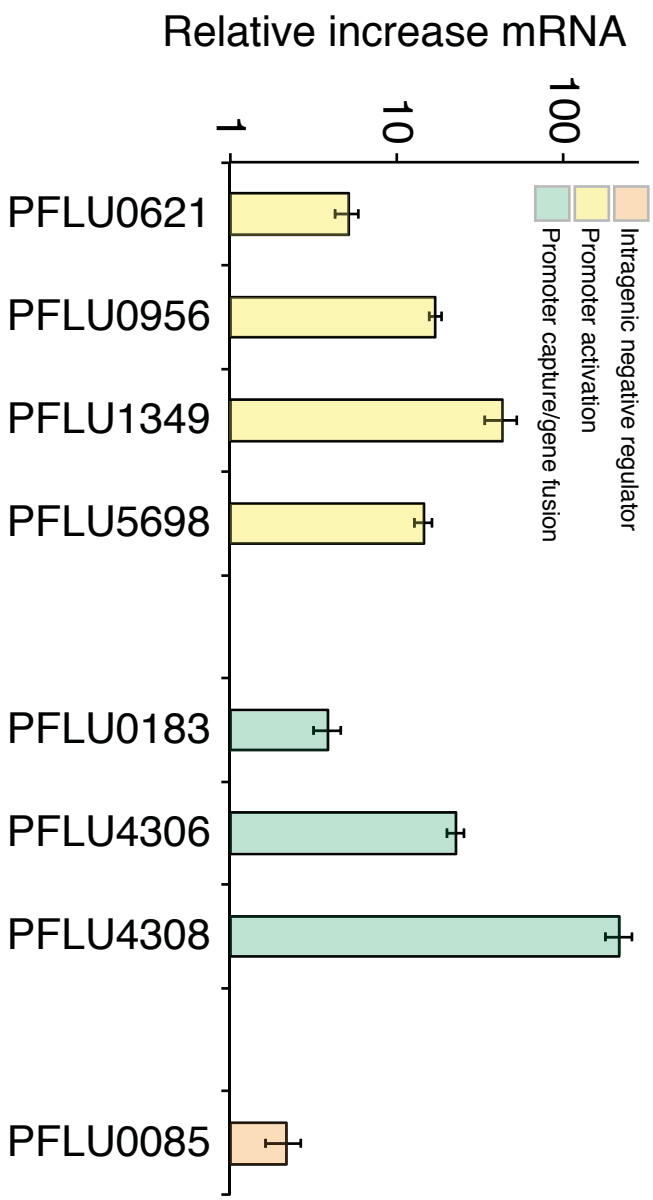

A
?
?
?
?
B
?
?
?

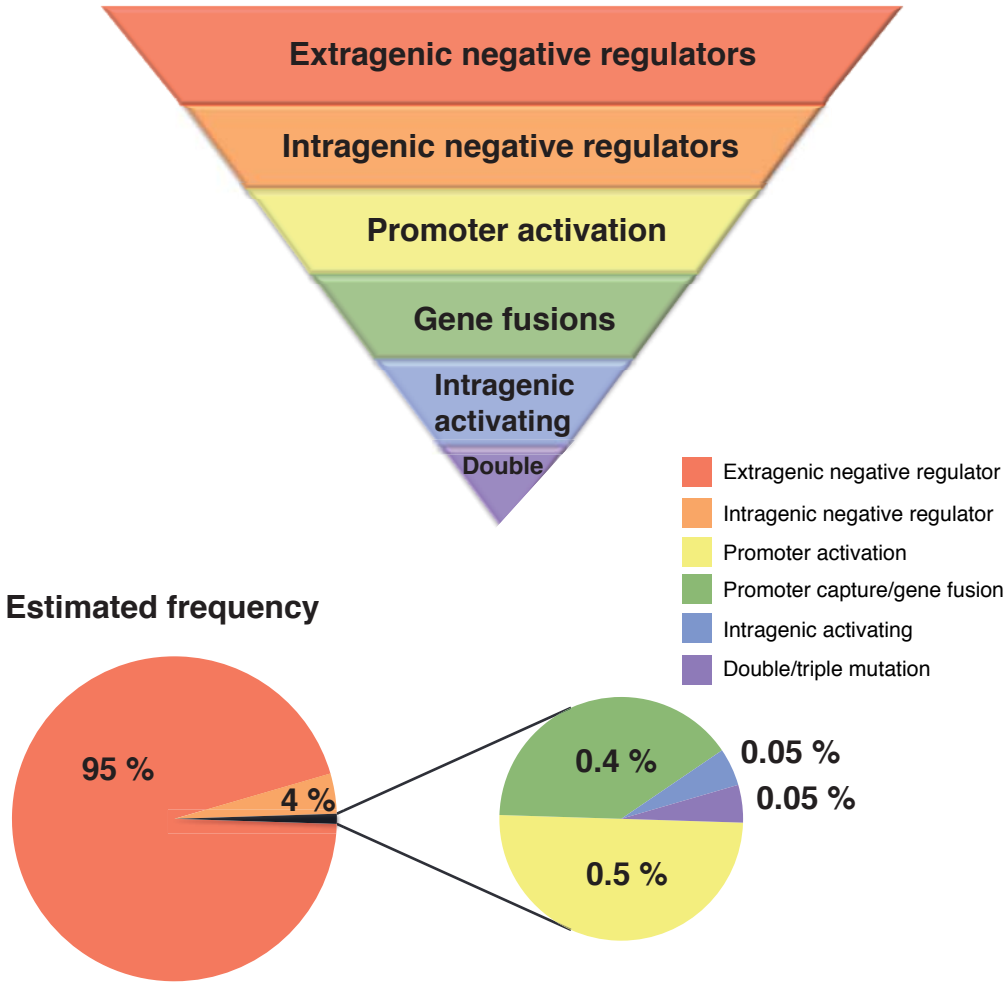

A
SM
Selection for access to oxygen,
colonization of air-liquid interface
WS
B
Loss-of-function mutation in
negative regulator
(WspF, AwsX, MwsR)
36 DGCs ?
Remove common pathways
and re experiment with
SM Δwsp Δaws Δmws
DGC
(WspR, AwsR, MwsR)
c-di-GMP
cellulose
(wss)
WS

Extragenic negative regulator		wspF*, awsX*
Intragenic negative regulator		mwsR*, PFLU0085 (43)
Promoter activation		PFLU0956 (9), PFLU5698 (7) PFLU1349 (2)
Promoter capture / gene fusion		PFLU0183 (8), PFLU4306 (8) PFLU4308 (1)
Intragenic activating		PFLU0956 (1), PFLU3448 (2), PFLU3571 (3), PFLU5960 (2)
Two extragenic negative regulators		PFLU0458, PFLU4744 (3)
Extragenic negative regulator and promoter activating		PFLU0458, PFLU0621 (1)
Three extragenic negative regulators		PFLU4414, PFLU4443, PFLU4744 (1)


A
Extragenic negative regulator
Intragenic negative regulator
Promoter activation
Promoter capture/gene fusion
Intragenic activating
Selection coefficient relative SBW25 SM
0.8
0.6
0.4
0.2
0
SM control
PFLU5960 D160G
PFLU5698 Promoter
PFLU0085 V447G
PFLU0085 L408P
PFLU4308 Fusion
PFLU3448 A200V
PFLU0085 Δ437-458
PFLU4306 Fusion
PFLU0085 Δ361-414
PFLU0183 Fusion
PFLU3571 W13R
PFLU0956 Promoter
MwsR V728A
PFLU1349 Promoter
WspF S301R
AwsX Δ77-87
B
Selection coefficient relative WspF S301R
0.1
0
-0.1
-0.2
-0.3
WspF control
PFLU0085 Δ361-414
PFLU0183 Fusion
AwsX Δ77-87
PFLU0085 V447G
PFLU0085 Δ437-458
PFLU5960 D160G
PFLU0085 L408P
PFLU4306 Fusion
PFLU4308 Fusion
PFLU3448 A200V
PFLU5698 Promoter
PFLU0956 Promoter
PFLU3571 W13R
PFLU1349 Promoter
MwsR V728A


A
Extragenic negative regulator
Promoter activation
Double/triple mutation
Control
PFLU0458 E45*
PFLU4744 S39N
Double
PFLU0458 I835S
PFLU0621 Promoter
Double
PFLU4443 E398*
PFLU4414 fs
PFLU4744 I44T
Double (4443 4414)
Double (4443 4744)
Double (4414 4744)
Triple
Selection coefficient relative SBW25 SM
0.8
0.6
0.4
0.2
0
B
Selection coefficient relative WspF S301R
0
-0.1
-0.2
-0.3
-0.4
-0.5


Relative increase mRNA
1
10
100
PFLU0621
PFLU0956
PFLU1349
PFLU5698
PFLU0183
PFLU4306
PFLU4308
PFLU0085
Promoter capture/gene fusion
Promoter activation
Intragenic negative regulator


Extragenic negative regulators
Intragenic negative regulators
Promoter activation
Gene fusions
Intragenic activating
Double
Extragenic negative regulator
Intragenic negative regulator
Promoter activation
Promoter capture/gene fusion
Intragenic activating
Double/triple mutation
Estimated frequency
95 %
4 %
0.4 %
0.05 %
0.05 %
0.5 %