2-12-2016

Adaptive evolution by spontaneous domain fusion and protein relocalisation

Andrew D Farr[1] and Paul B Rainey[1,2,3]

[1]New Zealand Institute for Advanced Study, Massey University, Auckland, 0745, New Zealand

[2]Department of Microbial Population Biology, Max Planck Institute for Evolutionary Biology, Plön 24306, Germany

[3]Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI ParisTech), CNRS UMR 8231, PSL Research University, 75231 Paris Cedex 05, France

Key words: gene fusion, relocalisation, diguanylate cyclase, cyclic-di-GMP, experimental evolution

Author contributions: ADF: Conception and design, acquisition of data, analysis and interpretation of data, drafting and revising the article. PBR: Conception and design, analysis and interpretation of data, drafting and revising the article.

Abstract

Knowledge of adaptive processes encompasses understanding of the emergence of new genes. Computational analyses of genomes suggest that new genes can arise by domain swapping, however, empirical evidence has been lacking. Here we describe a set of nine independent deletion mutations that arose during the course of selection experiments with the bacterium *Pseudomonas fluorescens* in which the membrane-spanning domain of a fatty acid desaturase became translationally fused to a cytosolic di-guanylate cyclase (DGC) generating an adaptive phenotype. Detailed genetic analysis of one chimeric fusion protein showed that the DGC domain had become membrane-localised resulting in a new biological function. The relative ease by which this new gene arose along with its profound functional and regulatory effects provides a glimpse of mutational events and their consequences that are likely to play a significant role in the evolution of new genes.

47 **Introduction**

48

49 The emergence of new genes by mutation – readily identified through comparative genomics
50 – provides an obvious and important source of adaptive phenotypes (Chen et al. 2013; Long
51 et al. 2013; Zhang and Long 2014). Mutational mechanisms involve divergence of duplicate
52 genes (Ohno 1970; Lynch and Conery 2000; Bergthorsson et al. 2007; Nasvall et al. 2012),
53 exon shuffling, the domestication of transposable elements, retrotransposition, gene fusion,
54 and the de novo evolution of new reading frames (Long et al. 2003; Kaessmann 2010; Tautz
55 and Domazet-Loso 2011; Ding et al. 2012; Long et al. 2013; Zhao et al. 2014).

56

57 Mutations that result in chimeric genes – by recombination of parental genes into a single
58 open reading frame, or by retrotransposition of a gene into an alternate reading frame – are
59 likely to generate new genes with relative ease (Rogers et al. 2010; Ranz and Parsch 2012).
60 This is because fusions stand to produce novel combinations of functional domains and
61 regulatory elements with few mutational steps. For example, promoter capture, whereby a
62 fusion event couples an existing gene to a new promoter, can cause abrupt changes in
63 temporal and spatial patterns of expression (Blount et al. 2012; Rogers and Hartl 2012;
64 Annala et al. 2013). Additionally, novel combination of domains may result in a range of post-
65 translational outcomes, ranging from relocalisation of domains, to novel inter-protein
66 associations, regulation of enzymatic activity and possibly even the formation of novel protein
67 functions (Patthy 2003; Bashton and Chothia 2007; Jin et al. 2009; Peisajovich et al. 2010;
68 Rogers and Hartl 2012; Singh et al. 2012)

69

70 Comparative computational analyses provide evidence for the evolutionary importance of
71 gene fusion events generating chimeric proteins. Compelling data comes a diverse range of
72 organisms, including bacterial species (Pasek et al. 2006) *Oryza sativa* (Wang et al. 2006),
73 *Drosophila* spp. (Long and Langley 1993; Wang et al. 2000; Wang et al. 2002; Jones et al.
74 2005; Rogers et al. 2009; Rogers and Hartl 2012), *Danio rerio* (Fu et al. 2010),
75 *Caenorhabditis elegans* (Katju and Lynch 2006) and humans (Courseaux and Nahon 2001;
76 Zhang et al. 2009). Despite the apparent importance of gene fusion events, the functional

77 effects of presumed instances of domain swapping have received little attention. Notable
78 exceptions are studies of "new" *Drosophila* genes: *sphinx*, *jingwei* and *Sdic*, which affect
79 courting behaviour, alcohol-dehydrogenase substrate specificity and sperm-motility,
80 respectively (Zhang et al. 2004; Dai et al. 2008; Zhang et al. 2010; Yeh et al. 2012). These
81 three genes have complex mutational histories, but originate from gene fusion events:
82 retrotransposition of distal genes (Long et al. 1999; Wang et al. 2002) or duplication and
83 fusion of neighbouring genes (Ranz et al. 2003).
84
85 Beyond studies of new genes inferred from computational analyses, synthetic biology has
86 shown that genes created by domain swapping can have important phenotypic effects. For
87 example, in vitro recombination of genes involved in yeast mating was shown to generate
88 greater diversity compared with control manipulations in which the same genes were
89 duplicated in the absence of recombination (Peisajovich et al, 2010). While an elegant
90 demonstration of the potential importance of gene fusion events in evolution, observation of
91 such events in natural populations is desirable.
92
93 Evolution experiments with microbes provide as yet unrealised opportunities to understand
94 evolutionary process, including mechanistic details. During the course of studies designed to
95 elucidate the range of mutational pathways to a particular adaptive "wrinkly spreader" (WS)
96 phenotype (Spiers et al. 2002; Goymer et al. 2006; Bantinaki et al. 2007; McDonald et al.
97 2009), we discovered a number of gene fusion events (Lind et al. 2015). Two classes bore
98 the hallmarks of promoter capture whereby deletions caused a focal gene to come under
99 control of a more highly expressed promoter eliciting the adaptive phenotype (Lind et al.
100 2015). A third class of mutation appeared not to conform to expectations of promoter
101 capture. Intriguingly, this class was defined by eight independent deletions each of which
102 caused a translational fusion between two adjacent genes. As we show here, these fusions
103 define new genes that are chimeras between a membrane-localised fatty acid desaturase and
104 a cytosolic di-guanylate cyclase (DGC).

Results

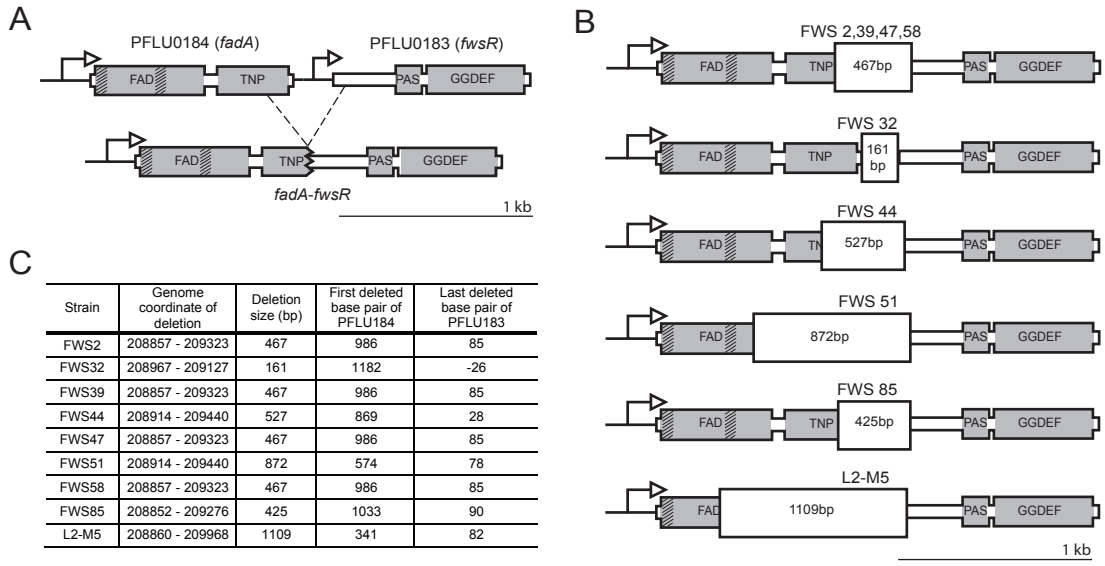
The WS model system and FWS types

When propagated in a spatially structured microcosm *P. fluorescens* SBW25 undergoes rapid diversification, producing a range of niche specialist types (Rainey and Travisano 1998). Among the best studied is the wrinkly spreader (WS) (Rainey and Travisano 1998; Spiers et al. 2002; Spiers et al. 2003; Goymer et al. 2006; Bantinaki et al. 2007; McDonald et al. 2009; Lind et al. 2015). WS types are caused by mutations in one of numerous genes that regulate or encode di-guanylate cyclases (DGCs). These genes catalyse the synthesis of 3´,5´-cyclic-di-guanosine monophosphate (c-di-GMP), a signalling molecule that allosterically regulates a complex of enzymes that produce an acetylated cellulosic polymer (ACP) (Ross et al. 1987; Tal et al. 1998; Amikam and Galperin 2006; De et al. 2008). Over-production of the polymer is the proximate cause of the WS phenotype (Spiers et al. 2002; Spiers et al. 2003).

Mutations that cause the WS phenotype in *P. fluorescens* SBW25 typically reside in one of three DGC encoding pathways: Wsp, Aws or Mws (McDonald et al. 2009). These pathways encode post-translational negative regulators that, when the targeted by loss-of-function mutations, result in constitutive DGC activity (Goymer et al. 2006). Elimination of these three pathways by deletion led to the discovery of a further 13 mutational routes to the WS phenotype, all involving pathways encoding DGC domains (Lind et al. 2015). Within this set of 12 new mutational routes were three loci where deletion events led to gene fusions. In two instances, the phenotypic effects were explained by increased transcription of DGCs resulting from promoter capture (Lind et al. 2015). However, for the third, which involved fusion between two open reading frames (*pflu0183* and *pflu0184*), there was little to suggest that the phenotypic effects derived from promoter capture (Lind et al. 2015).

All mutations involving *pflu0183* described in Lind *et al.* (2015) involved deletions spanning parts of *pflu0183* and *pflu0184*. Each deletion resulted in a predicted single open reading frame, transcribed from the promoter of *pflu0184* (Figure 1A), thus maintaining the reading frame of *pflu0183*. Of the eight independent mutations at this locus, four consisted of an identical 467bp deletion (see Figure 1B and 1C). A ninth fusion, “L2-M5”, obtained from an

134 independent experiment (see below) is also depicted in Figure 1. Reconstruction of the 467bp
135 deletion in the ancestral SM genotype was sufficient to cause the WS colony morphology,
136 niche specialisation and over-production of ACP (Figure 1 – figure supplement 1). All fusion
137 mutations involving *pflu0183* produced the characteristic WS phenotype: hereafter these are
138 referred to as "FWS" types and the focal FWS type studied here is termed "FWS2" (Figure 1 –
139 figure supplement 2).
140



C

Strain	Genome coordinate of deletion	Deletion size (bp)	First deleted base pair of PFLU184	Last deleted base pair of PFLU183
FWS2	208857 - 209323	467	986	85
FWS32	208967 - 209127	161	1182	-26
FWS39	208857 - 209323	467	986	85
FWS44	208914 - 209440	527	869	28
FWS47	208857 - 209323	467	986	85
FWS51	208914 - 209440	872	574	78
FWS58	208857 - 209323	467	986	85
FWS85	208852 - 209276	425	1033	90
L2-M5	208860 - 209968	1109	341	82

141

142 **Figure 1: Arrangement of *pflu0184* and *pflu0183* in the ancestral SM genotype and following**
143 **gene fusion**. **A)** Illustrated is the fusion mutation that occurred on four independent occasions
144 (underpinning FWS 2, 39, 47 and 58). Grey areas represent predicted domains and shading within grey
145 areas depict predicted transmembrane domains. *Pflu0184* (referred to as *fadA*) consists of a FAD region
146 encoding a predicted fatty acid desaturase at residues 11 to 246. 'TNP' is a predicted transposase
147 element (DDE_Tnp_ISL3 family) at residues 261 to 389. The neighbouring *pflu0183* (referred to as
148 *fwsR*) encodes a predicted PAS fold (of the PAS_3 family) at residues 68 to 157, and a predicted
149 GGDEF domain at residues 174 to 331. The putative FadA protein contains two transmembrane
150 domains (TMDs) between residues 10 to 32 and 135 to 157 as predicted by the 'TMHMM server v2.0'
151 (Krogh et al. 2001). Dashed lines represent the deletion that defines FWS 2, 39, 47 and 58. **B)**
152 Illustration of the nine independent *fadA-fwsR* fusions at *fadA* and *fwsR*. The mutations in FWS 2, 39,
153 47 and 58 share a common deletion junction. All deletions conserved the predicted ancestral reading
154 frame for the GGEEF domain. **C)** Genomic details of *fadA-fwsR* fusion mutations. Positions are relative
155 to the first base pair of the predicted open reading frame of *pflu0184* or *pflu0183*, or to the
156 *Pseudomonas fluorescens* SBW25 genome (Silby et al. 2009).

157 **The *fadA-fwsR* fusion**

158 *Pflu0183* encodes a predicted di-guanylate cyclase (DGC) henceforth referred to as '*fwsR*'.
159 FwsR forms a predicted protein of 335 residues in length with a predicted PAS fold and a
160 GGDEF domain at the *C*-terminus (Figure 1A). This GGDEF domain features a GGEEF motif,
161 which is indicative of DGC activity (Chan et al. 2004; Galperin 2004; Malone et al. 2007;
162 Wassmann et al. 2007). The neighbouring gene *pflu0184* encodes a predicted protein 394
163 residues in length, including a predicted *N*-terminal fatty acid desaturase (residues 11 and
164 246) and a predicted transposase element (residues 261 to 289). *Pflu0184* is hereafter
165 referred to as '*fadA'*. FadA contains two predicted transmembrane domains (TMDs) between
166 residues 10 to 32 and 135 to 157. The gene arising following deletion is termed '*fadA-fwsR*'.
167
168 Little is known about the functions of proteins encoded by either *fadA* or *fwsR*. FadA shares
169 83% identity with DesA (PA0286) from *Pseudomonas aeruginosa* (Winsor et al. 2011), which
170 is a fatty acid desaturase (Zhu et al. 2006). Fatty acid desaturases modify phospholipids in
171 the cell membrane in order to modify membrane fluidity in response to environmental change
172 (Zhang and Rock 2008). In *P. aeruginosa*, transcription of *desA* is promoted during anaerobic
173 conditions resulting in increased membrane fluidity as a consequence of double bonds
174 introduced at the 9-position of fatty acid acyl chains (Zhu et al. 2006). Increased production of
175 short-chained fatty acids has recently been shown to increase production of cyclic di-GMP by
176 the Wsp pathway (Blanka et al. 2015). *fwsR* has orthologs across members of the genus
177 *Pseudomonas*, however, none of these have been the subject of study.
178
179 It is not known whether the proximal relationship of *fadA* and *fwsR* orthologs across many
180 *Pseudomonas* spp. reflects a functional or regulatory relationship between the two genes (or
181 their protein products), however the operon prediction tool *DOOR* suggests they are separate
182 transcriptional units (Mao et al. 2014). In species such as *P. putida* and *P. stutzeri* orthologs
183 of *fadA* and *fwsR* are adjacent to each other (Figure 1 – figure supplement 3). However, in *P.*
184 *aeruginosa* the locus of the GGDEF domain-encoding gene (PA0260) is separated from *desA*
185 by three open reading frames (ORFs).
186

DGC activity of FadA-FwsR is necessary to generate FWS2

How do deletions between *fadA* and *fwsR* (and generation of *fadA-fwsR*) produce FWS phenotypes? Like other WS-causing mutations (Goymer et al. 2006; Bantinaki et al. 2007; McDonald et al. 2009) the causal *fadA-fwsR* fusion is likely to alter the activity of the DGC domain of *fwsR* resulting in over-production of c-di-GMP. The constrained length of the spontaneous deletion mutations causing FWS types (all *fadA-fwsR* fusions arose by deletion of no more than 90 bp of *fwsR* – see Figure 1C) suggests that activity of the *fwsR*-encoded DGC is required for the FWS phenotype. To test this hypothesis the conserved GGEEF motif that defines the active site of the DGC was replaced by a mutant allele (GGAAF) expected to eliminate DGC function (Malone et al. 2007). The phenotype of FWS2 carrying the mutant *fadA-fwsR* allele was smooth, non-cellulose (ACP) producing and unable to colonise the air-liquid interface of static broth microcosms (Figure 2). This demonstrates that diguanylate cyclase activity of *fadA-fwsR* is necessary for the FWS2 phenotype.

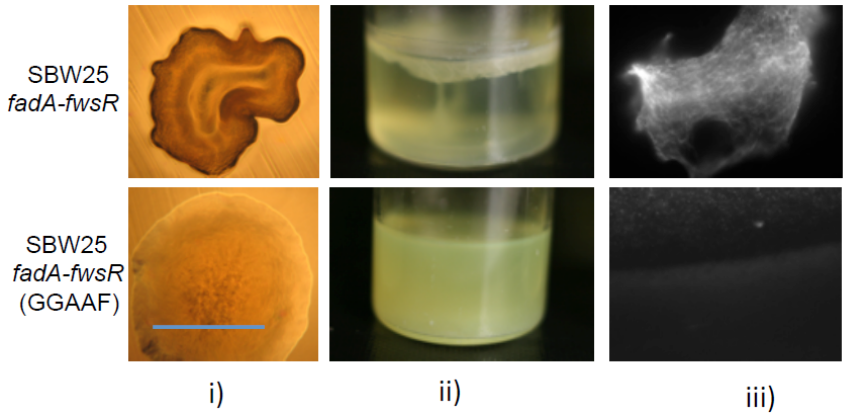
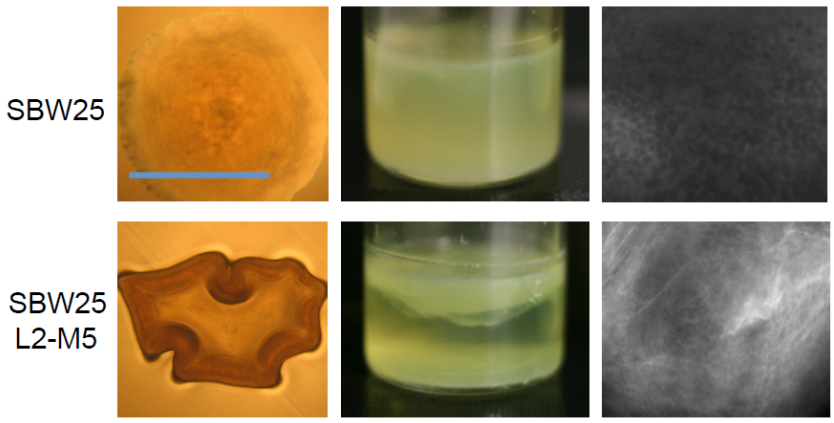


Figure 2: DGC activity encoded by *fadA-fwsR* is necessary to cause WS. In the context of *fadA-fwsR* reconstructed in SBW25, substitution of a GGAAF motif for the ancestral GGEEF motif within the predicted active site of the DGC domain results in SM phenotypes in i) morphology (visualised by light microscopy (10x objective) of ~32 h colonies grown on King's B agar (KBA), blue bar is ~2mm); ii) colonised niche (depicted microcosms were inoculated with single colonies incubated statically at 28°C for 3 days)); and iii) ACP production, as indicted by calcofluor fluorescence (visualised by fluorescent microscopy, 60x objective).

211 **FWS2 does not require fatty acid desaturase activity**

212 A *fadA-fwsR* fusion mutant isolated during the course of an independent experiment, termed
213 'L2-M5', contains a *fadA-fwsR* deletion that removed 347 bp of the predicted FadA domain
214 (Figure 1B and 1C). This deletion results in a fusion of nucleotide 340 of *fadA* to nucleotide 83
215 of *fwsR*, while preserving the reading frame of *fwsR*. This 'L2-M5' mutation resulted in a
216 characteristic WS phenotype upon substitution for *fadA* and *fwsR* in the ancestral SM
217 genotype (Figure 3). That a *fadA-fwsR* fusion causes FWS, despite missing most of the
218 enzymatic domain shows that fatty acid desaturase activity is not required for the FWS
219 phenotype. It also suggests that structural componentry (such as non-enzymatic domain
220 folds) of FadA is not required to cause FWS, unless that structure is proximal to the *N*-
221 terminus.
222



223

224 **Figure 3: The fatty acid desaturase domain is not required for the phenotypic effects of *fadA-***
225 ***fwsR*.** The phenotype of SBW25 expressing the L2-M5 *fadA-fwsR* fusion generates the characteristic
226 WS phenotype, niche preference, and produces ACP as evidenced by presence of a fluorescent signal
227 after staining a portion of the colony with calcofluor (see Figure 2 for details of each image).

228

229 **Altered transcription of *fwsR* is insufficient to cause FWS**

230 One hypothesis to account for the effects of the *fadA-fwsR* fusion predicts effects mediated
231 via transcription. Promoter mutations at several GGEEF domain-encoding genes have been
232 associated with transcriptional increases and the evolution of WS types (Lind et al. 2015). In
233 the case of the *fadA-fwsR* fusion, this possibility is given credence by the presence of a rho-
234 independent transcriptional terminator downstream of the stop-codon of *fadA* (predicted by

the WebGeSTer transcription terminator database (Mitra et al. 2011)). Removal of this terminator by the spontaneous deletion mutation could lead to increased transcription of *fwsR,* which may in turn elevate intracellular levels of the DGC, resulting in increased production of c-di-GMP and ultimately the FWS phenotype.

To test this hypothesis the promoter of *fadA* was fused to *fwsR* (resulting in P_{fadA}-*fwsR,* Figure 4A) and integrated into the genome such that it replaced the *fadA fwsR* locus in the ancestral SM genotype. No effect of the *fadA* promoter fusion was observed: the genotype retained its smooth colony morphology, did not colonize the air-liquid interface, or produce ACP (Figure 4B). This demonstrates that changes in *fwsR* transcription caused by the *fadA-fwsR* fusion are not sufficient to cause WS.

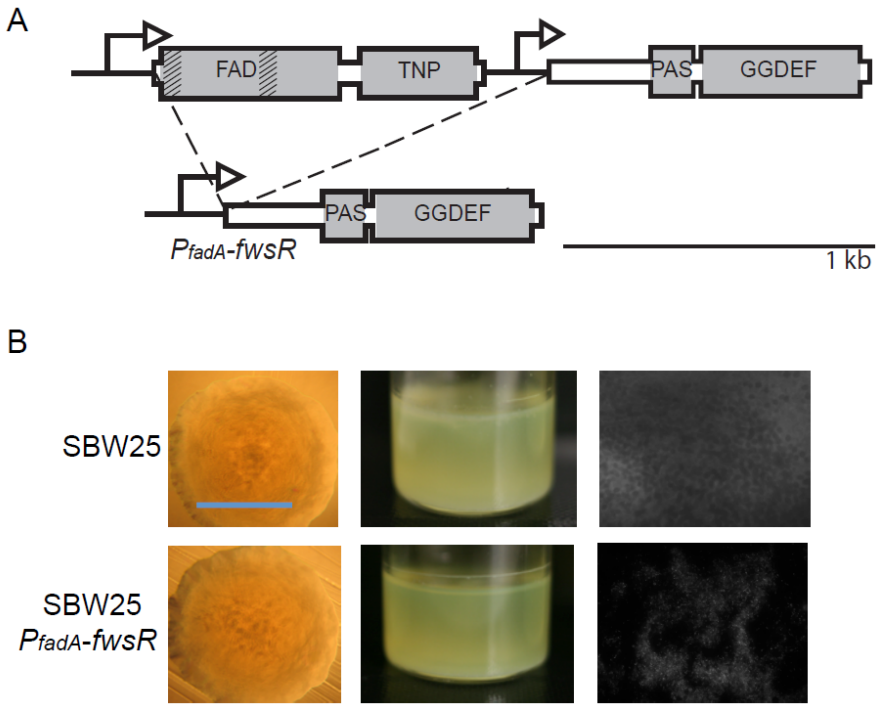


Figure 4: Altered transcription of *fwsR* is insufficient to cause FWS. A) Illustration of construct P_{fadA}-*fwsR*. In this construct the promoter for *fadA* is fused to *fwsR*. **B)** The potential change in transcription caused by promoter capture does not cause the WS phenotype. The phenotype of the P_{fadA}-*fwsR* construct reconstructed in the ancestral SM genotype is phenotypically SM in terms of morphology, niche specialization and absence of ACP production (see Figure 2 for details of each image).

Translational fusion of FadA to FwsR is necessary for FWS2

The alternate hypothesis, namely, that the translation fusion is itself necessary to generate FWS2 was next tested. This was achieved by construction of a *fadA-fwsR* fusion in which translational coupling was eliminated: 182 bp between the stop codon of *fadA* and the ATG start codon of *fwsR* were deleted and replaced by stop codons (one in each reading frame). The *in situ* ribosome-binding site (RBS) of *fwsR* was left intact to allow translation initiation of an independent protein product for *fwsR*. The allele was introduced into the FWS2 background (where it replaced the *fadA-fwsR* fusion of FWS2) generating *fadA-3X-fwsR* (Figure 5A). The phenotype of this mutant resembled the ancestral SM type in all respects (Figure 5B). This indicates that translational fusion between *fadA* and *fwsR* is necessary to generate FWS2.

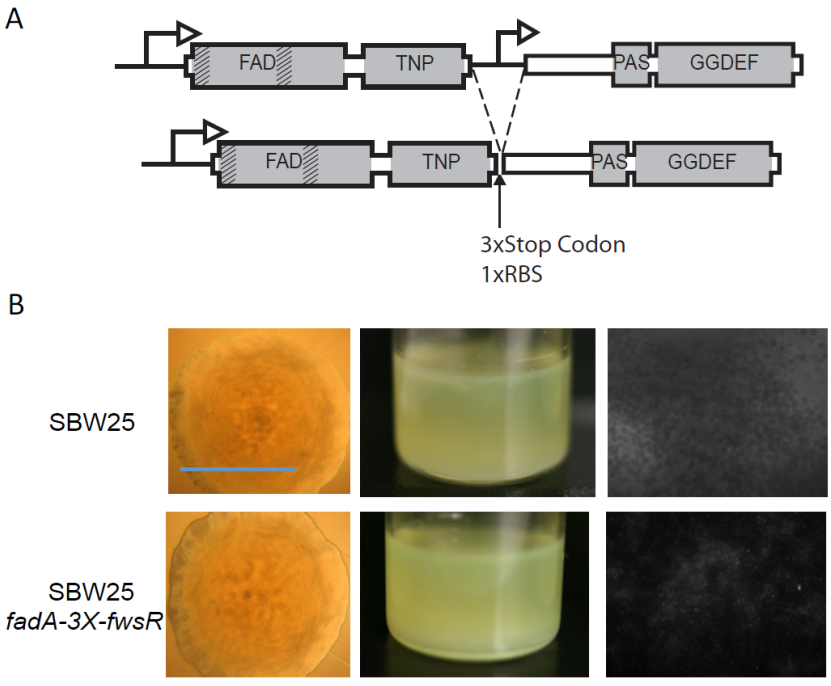


Figure 5: The FWS phenotype requires translational fusion of the *fadA* and *fwsR* genes. A) Illustration of the construct *fadA-3X-fwsR*. The *fadA* open reading frame was fused to that of *fwsR* and is separated by three stop codons. The ancestral ribosome-binding site (RBS) of *fwsR* was retained. **B)** Transcriptional fusion and translational uncoupling of *fadA* and *fwsR* do not cause FWS. The phenotype of *fadA-3X-fwsR* reconstructed in the ancestral SM genotype does not cause the FWS phenotype, indicating translational fusion of the genes is required to cause FWS. The *fadA-3X-fwsR* construct causes a characteristic SM phenotype in morphology, broth colonisation and lack of ACP production (see Figure 2 for details of each image).

The FadA-FwsR fusion relocates the DGC domain

To test the hypothesis that *fadA-fwsR* results in localisation of FwsR to the membrane, the cellular locations of proteins encoded by *fadA*, *fwsR* and the *fadA-fwsR* fusion were visualised by creating a green fluorescent protein (GFP) translational fusion to the *C*-terminal region encoded by each gene. All fusions were cloned into the multiple cloning site (MCS) of mini-Tn*7*-*lac* and modified to contain a *Pseudomonas*-specific RBS (attempts to visualise foci using native gene promoters proved unsuccessful). Two controls were used during microscopy: a negative control of the ancestral SM genotype without a *gfp* insert in the mini-Tn7-*lac* cassette; and a positive control of the ancestral SM type containing *gfp* expressed in the integrated mini-Tn7-*lac* cassette. A cassette expressing *fadA-3X-fwsR-gfp* was also made to confirm that the *fadA-3X-fwsR* construct allowed independent translation of FwsR.

The fluorescent signal from each genotype was quantified to show that non-specific fluorescent signals from FwsR-GFP and FadA-3X-FwsR-GFP are significantly greater than the negative control (*Welch t*-test, *p*-value = 4.8×10^{-10} and 5.0×10^{-8}), indicating that the diffuse expression of fluorescence results from the induced expression of GFP fusions (Figure 6).

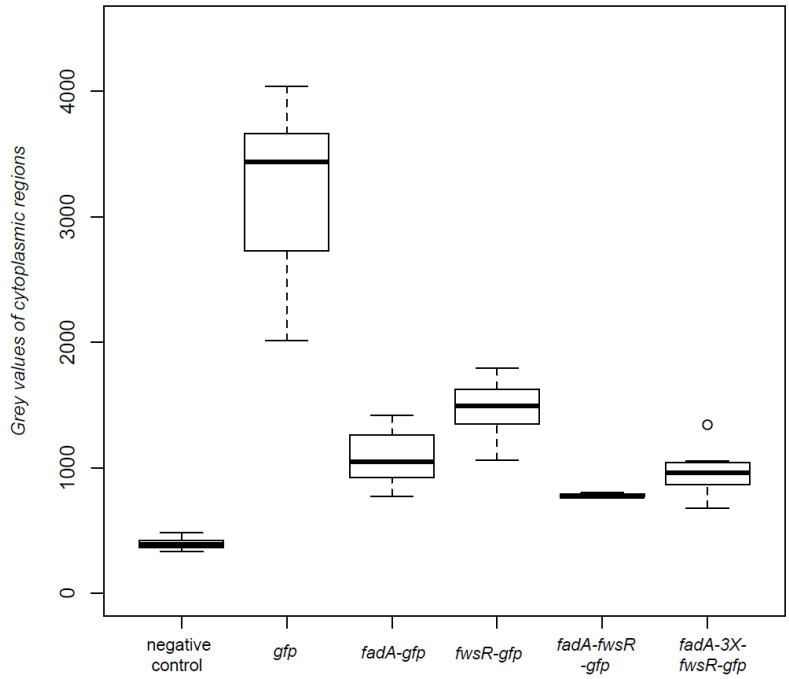


Figure 6: Cytoplasmic levels of fluorescence from protein fusions are above background levels. Depicted is a box plot of the florescent signals resulting from *gfp* translational fusions to fadA, fwsR and *fadA-fwsR* fusion variants. The mean fluorescent signal of 12 randomly picked cells (4 cells per independent culture) was obtained by making lateral transects of cells observed under fluorescent microscopy. The mean fluorescent signal of each cell was normalised to the background fluorescence of each image. Increases in the levels of fluorescence from the negative control demonstrate induced expression of genes fused with GFP. Data is presented as standard boxplots. Dark bars represent median values, boxes are the range from first to third quartile, whiskers represent minima and maxima, and circles represent values outside the 1.5 interquartile range.

Fluorescence microscopy demonstrates that the *fadA-fwsR* fusion alters the location of the FwsR DGC from the cytosol to the membrane (Figure 7). The location of the fluorescent signal of cells expressing *fwsR-gfp* is visually diffuse and is located within the cytoplasm (Figure 7C). Similarly, the signal from *fadA-3X-fwsR-gfp* is dispersed throughout the cytoplasm (Figure 7D). In contrast, cells expressing *fadA-gfp* have clear foci localised to the edge of the cell and, by inference, near the membrane (Figure 7E). The foci are distributed predominately at the lateral edge of cells in visual reference to the phase contrast images, similar to observations of the localisation of WspA in *P. aeruginosa* (O'Connor et al. 2012). The protein expressed by *fadA-fwsR-gfp* (Figure 7F) is localized in a manner identical to that

observed in cells expressing *fadA-gfp*. The visual co-localisation of foci with the edge of the cells in genotypes expressing *fadA-gfp* and *fadA-fwsR-gfp* was confirmed by co-localisation analysis by using Van Steensel's approach (van Steensel et al. 1996) (Figure 7 – figure supplement 1). Together, the location of fluorescent foci seen in cells with induced *fadA-gfp* and *fadA-fwsR-gfp* demonstrates the *fadA-fwsR* fusion event relocated FwsR from the native cytosol to the membrane location of FadA.

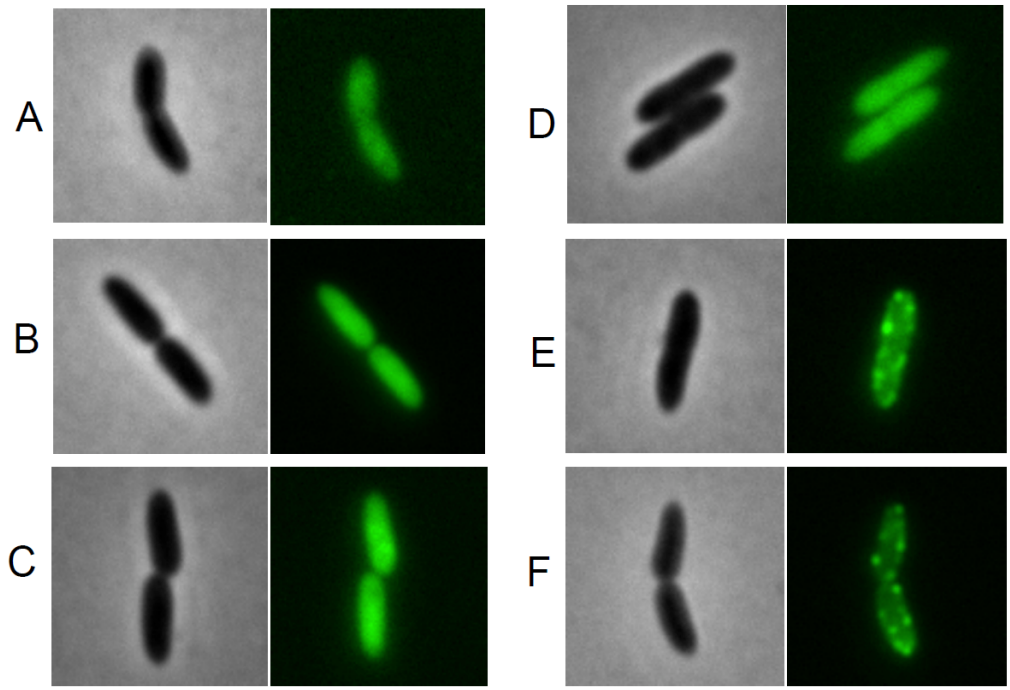


Figure 7: Fluorescence microscopy depicting the distribution of GFP tagged proteins encoded from the *fwsR* locus. Representative phase contrast (left) and GFP fluorescent images (right) of the subcellular distribution of fluorescence of A) *gfp*(-) control (showing background autofluorescence); B) *gfp*(+) control; C) *fwsR-gfp*; D) *fadA-3X-fwsR-gfp*; E) *fadA-gfp* and F) *fadA-fwsR-gfp*. Genotypes were grown in minimal media overnight and subcultured in minimal media with 1 mM IPTG for approximately 2 hours.

Localisation of FwsR to the membrane is sufficient to generate FWS

If membrane localisation of FwsR is all that is required to activate synthesis of c-di-GMP, then fusion of FwsR to any membrane spanning domain-containing protein should suffice to generate the FWS phenotype. To test this hypothesis, the membrane-spanning domain of *mwsR* (PFLU5329) was fused to *fwsR*. Details of the construct are shown in Figure 8A.

Replacement of the native *fadA fwsR* locus with the *mwsR-fwsR* fusion in the ancestral genotype lacking *mwsR* (to avoid unwanted allelic exchange at this gene) resulted in the FWS phenotype (Figure 8B).

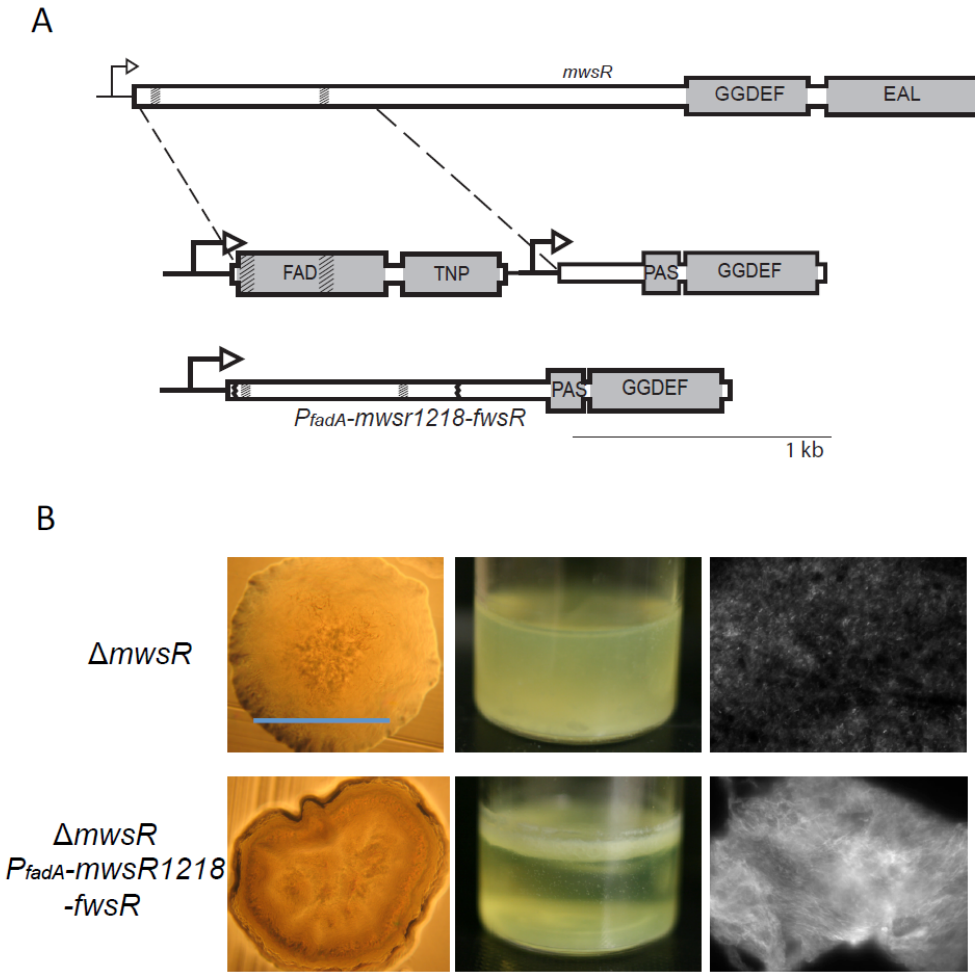


Figure 8: The translational fusion of the transmembrane regions of *mwsR* to *fwsR* causes the FWS phenotype. A) Illustration of the P_{fadA}-*mwsR1218-fwsR* construct. The first 1218 bp of *mwsR* (which encodes for two TMD regions) was used to replace the *fadA* element of *fadA-fwsR*. Vertical grey bands within the ORFs represent TMDs. **B)** The phenotype of P_{fadA}-*mwsR1218-fwsR* is FWS. The P_{fadA}-*mwsR1218-fwsR* construct reconstructed in an SM ancestral background is characteristically WS in morphology, niche habitation and ACP production (see Figure 2 for details of each image).

346 **Discussion**

347

348 The origin of new genes is a subject of fundamental importance and longstanding debate
349 (Sturtevant 1925; Haldane 1932; Bridges 1936). Gene duplication and divergence, once seen
350 as the primary source (Ohno 1970; Lynch and Conery 2000; Nasvall et al. 2012), is now
351 recognized as just one of a number of routes by which new genes are born. Studies in
352 comparative genomics indicate that new genes have arisen from retroduplication (in which
353 mRNA is reverse transcribed into a complementary DNA copy and inserted into the
354 chromosome (Brosius 1991; Long et al. 2003) from retrotransposon-mediated transduction
355 (Moran et al. 1999; Cordaux and Batzer 2009), from deletion and recombination events that
356 generate chimeric gene fusions (Marsh and Teichmann 2010; Ranz and Parsch 2012) from
357 genomic parasites (Volff 2006; Feschotte and Pritham 2007), and even from previously non-
358 coding DNA (Tautz & Domazet-Loso 2011, Neme & Tautz 2013).

359

360 Despite the apparent range of opportunities for the birth of new genes, there are few
361 examples in which the evolution of a new gene has been captured in real time, the selective
362 events leading to its formation understood, and mechanistic details underpinning function of
363 the new gene revealed. Those thus far reported have involved recombination or deletion
364 events leading to promoter capture. For example, in the Lenski long-term evolution
365 experiment, Blount et al (2012) described a rare promoter capture event that underpinned
366 evolution of citrate utilization in *E. coli*. Similarly, Lind et al (2015), using experimental
367 *Pseudomonas* populations reported promoter capture events caused by deletions that
368 increased transcription of genes encoding active DGCs necessary for over production of c-di-
369 GMP and the evolution of WS types.

370

371 Here, provided with opportunities afforded by experimental evolution, we have observed, in
372 real time, multiple independent deletion events, each of which caused a translational fusion
373 between two genes: the membrane-spanning domain of *fadA* and the DGC domain of *fwsR*.
374 Each fusion resulted in the birth of a new gene with the resulting fusion altering the cellular
375 location of the DGC domain. This alteration conferred new biological properties: activation of

the DGC, the synthesis of c-di-GMP, over-production of cellulose and generation of the adaptive wrinkly spreader phenotype.

In many respects, these events mirror those inferred from comparative studies responsible for new gene function, for example, the human gene *Kua-UEV* is thought to have originated from fusion of *Kua* and *UEV* resulting in localization of UEV to endomembranes by virtue of the *N*-terminal localization domain of *Kua* (Thomson et al. 2000). That similar, albeit simpler, events were detected in experimental bacterial populations is remarkable given that prokaryotic cell structure presents few opportunities for protein re-localization to occur; it thus adds weight to the suggestion that the birth of new genes via fusions that re-localize proteins is likely to be more common than recognized (Buljan et al. 2010). Indeed, inferences from comparative genomics support this prediction (Byun-McKay and Geeta 2007): approximately 24 - 37% of duplicated gene pairs in *Saccharomyces cerevisiae* encode proteins that locate to separate cellular compartments (Marques et al. 2008). Of human multi-gene families known to encode proteins predicted to locate to the mitochondria, approximately 64% contain a gene predicted to relocate to an alternative subcellular location (Wang et al. 2009). Additionally, studies on the fate of duplicated genes in *C. elegans* suggest that approximately a third of new genes caused by duplication mutations are chimeric (Katju and Lynch 2003; Katju and Lynch 2006). Such chimeric mutations can introduce spatial encoding motifs to the duplicated protein, providing ample opportunity for the relocalisation of protein domains.

While the de novo fusion events reported here occurred between two adjacent loci, the fact that the DGC domain of *fwsR* could be fused (in vitro) to the membrane spanning domain of a gene more than one million nucleotides away (*mwsR*) and with similar effect reveals the considerable potential to generate new gene functions via fusion of distal genetic elements. This realization warrants consideration in context of standard models for the origin of new genes via duplication (amplification) and divergence (Ohno 1970, Bergthorsson et al 2007, Lynch and Conery 2000, Nävsall et al 2012). Typically divergence is considered a gradual process occurring via point mutations and aided by selection for original gene function plus selection for some promiscuous capacity (Bergthorsson et al 2007, Nävsall et al 2012).

406 However, formation of chimeric proteins by gene fusion (following duplication) provides
407 opportunity for divergence to occur rapidly and in a single – or small number – of steps. Such
408 distal-fusion events have been reported from genomic studies (Rogers et al. 2009, Rogers
409 and Hartl 2012) and shows how the modular nature of spatial localising domains can facilitate
410 rapid generation of new functions (Pawson & Nash 2003).
411
412 With few exceptions (see for example studies on WspR (Bantinaki et al. 2007; Malone et al.
413 2007; De et al. 2008; O'Connor et al. 2012; Huangyutitham et al. 2013) and PleD (Aldridge et
414 al. 2003; Paul et al. 2007) biochemical details underpinning the mechanisms of DGC
415 activation and factors determining the specificity of c-di-GMP signaling are yet to be fully
416 understood (Schirmer and Jenal 2009; Massie et al. 2012; Chou and Galperin 2016). The
417 mechanism by which the *fadA-fwsR* fusion leads to DGC activation is unknown. The
418 requirement for an active FwsR DGC domain and for that active domain to localize to the
419 membrane makes clear that membrane localization is a necessary condition. One possibility
420 is that membrane localization is alone sufficient to activate DGC activity (a suggestion
421 bolstered by the fact that the DGC domain of *fwsR* could be fused to the membrane-spanning
422 domain of *mwsR* with the same effect). For example, membrane localization may promote
423 homodimerization, which is necessary for DGC activity (Wassmann et al. 2007). Activation
424 may also be affected by fatty acid composition of the cell membrane as recently shown in *P.*
425 *aeruginosa* (Blanka et al. 2015). Alternatively, localization of FwsR to the membrane may
426 serve to facilitate spatial sequestration of c-di-GMP and thus molecular interactions between
427 the DGC and the membrane localized cellulose biosynthetic machinery (Hengge 2009).
428 Biochemical studies are required to explore these possibilities, but discovery that the DGC
429 domain of FwsR becomes active upon localization opens new opportunities for understanding
430 mechanisms of DGC activation and signaling, and points to the importance of connections
431 between DGCs and the cell membrane. In this regard it is significant that DGCs are
432 invariably membrane associated, or components of regulatory pathways that are – like Wsp
433 (Bantinaki et al. 2006)) – membrane localized.
434
435 One notable feature of genes that encode DGC domains is the diverse array of domains with

436 which they associate. Prior bioinformatic analysis of DGCs shows connections to numerous
437 domains, ranging from CheY-like domains to kinases and phosphodiesterases (Galperin
438 2004; Jenal and Malone 2006; Romling et al. 2013). Gene fusion events such as we have
439 described here are likely to contribute to the origin of these diverse spectra suggesting
440 capacity of DGC domains to be readily accommodated within proteins of diverse function. If,
441 as the evidence suggests, this is true, then questions arise as to why the underpinning
442 evolutionary events have not previously been observed in experimental studies of evolution
443 using bacteria. Our recent work provides a clue: of all single-step mutational routes to the
444 wrinkly spreader phenotype, fusions generating chimeric proteins constitute almost 10 % of
445 such events – but only when the genome is devoid of readily achievable loss-of-function
446 routes to the adaptive phenotype (if loss-of-function routes are intact then fusions described
447 here constitute ~0.1% of the total) (Lind et al. 2015). While such loss-of-function routes are
448 readily realised in laboratory populations where functional redundancy among DGC domains
449 is observed, the very real possibility is that in populations in natural environments such
450 redundancy does not exist, making loss-of-function routes less achievable and mutations of
451 the kind described here more likely. Taking experimental evolution into the wild – or at least
452 scenarios more reflective of the complex challenges faced by natural populations
453 (Hammerschmidt et al. 2014; Bailey and Bataillon 2016; Lind et al. 2016) is an important
454 future goal likely to shed new light on the origins of genes.
455
456
457
458
459
460
461
462

Materials and Methods

Media and strains

All *P. fluorescens* strains are derivatives of *P. fluorescens* SBW25 (Silby et al. 2009) and were cultured at 28°C in King's medium B (KB) (King et al. 1954). *Escherichia coli* DH5-α λpir (Hanahan 1983) was used to replicate all plasmids used to construct mutations. *E. coli* was cultured at 37°C in Lysogeny Broth (LB) (Bertani 1951). Solid media were prepared by the addition of 1.5% bacteriological agar. Strains intended for fluorescence microscopy were cultured in M9 minimal media (Sambrook and Russell 2001) containing 0.4% w/v glycerol and omitting thiamine. All bacterial overnight cultures were grown shaking at 160 rpm for approximately 16 hours in 30 mL glass microcosms containing 6 mL of medium. Antibiotics for the maintenance of plasmids were used at the following concentrations: gentamycin (Gm) 25 mg L^{-1}, kanamycin (Km) 100 mg L^{-1}, tetracycline (Tc) 15 mg L^{-1}, nitrofurantoin (NF) 100 mg L^{-1}, and cycloserine 800 mg L^{-1}. X-gal (5-bromo-4-chloro-3-indolyl-β-D- galactopyranoside) was used at a concentration of 60 mg L^{-1} in agar plates, IPTG at 1 mM in liquid culture and Calcofluor (fluorescent brightener 28) at a concentration of 35 mg L^{-1} in agar plates.

Mutation reconstruction

Strand overlap extension (SOE-PCR) was employed to construct all site-directed genomic mutations in *P. fluorescens* as previously described (Ho et al. 1989; Rainey 1999; Bantinaki et al. 2007). Briefly, Phusion High-Fidelity DNA polymerase (New England Biolabs) was used to generate an amplicon product approximately 1000 bp either side of the mutation, using overlapping primer sets with the required mutation within the 5' region of the internal primers. A second round of PCR using only the external primers was performed to make a single amplicon with the required mutation. The amplicon was then cloned into pCR8/GW/TOPO (Invitrogen) using TA cloning, transformed into an *E. coli* host and the plasmid was sequenced to confirm the presence of the mutation and no additional mutations within the amplified region. The DNA regions were then excised from pCR8/GW/TOPO by restriction digest and ligated into the pUIC3 suicide vector (Rainey 1999), and was then introduced into the *P. fluorescens* host by two-step allelic exchange (Kitten et al. 1998). This involved

493 transconjugation of pUIC3 into the *P. fluorescens* host by tri-parental mating, and selection for
494 transconjugants with the pUIC3 vector homologously recombined at the target sequence.
495 Overnight non-selective cultures of the transconjugates were then treated with tetracycline in
496 order to inhibit growth of cells that had lost the pUIC3 vector. The culture was then treated
497 with cycloserine to kill growing cells featuring the pUIC3 vector and thus enrich the fraction of
498 cells that were either isogenic with the host, or which now featured the mutation following a
499 second recombination event. The culture was serially diluted and plated on KB plates
500 containing X-gal to enable visual screening for genotypes that had lost the pUIC3 vector.
501 Several resulting white colonies were isolated and Sanger sequencing confirmed the
502 presence of the mutant allele and the absence of mutations that may have been introduced
503 during the cloning process. The sequence of each genomic mutation is accessible through
504 the following accession numbers: *fadA-fwsR* GGAAF (KU248756), *fadA-fwsR* L2M5
505 (KU248757), P_{fadA}-*fwsR* (KU248758), *fadA-3X-fwsR* (KU248759) and P_{fadA}-*mwsR1218-fwsR*
506 (KU248760).
507
508 The visualization of proteins required translational fusions of *fadA, fwsR* and fusions of these
509 genes with *gfp* variant *gfpmut3** (Andersen et al. 1998). The open reading frame of genes to
510 be tagged were cloned into pCR8 and then moved (by digestion with restriction
511 endonucleases and ligation) into the multiple cloning site (MCS) of mini-Tn*7*-LAC (Choi and
512 Schweizer 2006). This plasmid was modified to contain both a *Pseudomonas*-specific
513 ribosome-binding site (RBS) as well as a *gfp* encoding region 3' of the MCS. This resulted in
514 the ability to induce expression of our protein of interest, tagged at the *C*-terminus with GFP.
515 These plasmids were sequenced to identify possible mutations introduced during cloning.
516 Plasmids were then taken up by electrocompetent *P. fluorescens* SWB25 cells and the
517 integration of the construct at the *attB* site was confirmed by PCR and electrophoresis. The
518 sequence of each *gfp*-tagged construct cloned in the mini-Tn7-LAC MCS is accessible
519 through the following accession numbers: *gfp* positive control (KU248761), *fwsR-gfp*
520 (KU248762), *fadA-gf*p (KU248763), *fadA-fwsR-gfp* (KU248764) and *fadA-3X-fwsR-gfp*
521 (KU248765).
522

Microscopy techniques

Colony-level photography was performed using a Canon Powershot A640 camera in conjunction with a Zeiss Axiostar Plus light microscope using a 10x objective. Microscopy of cells was visualised using an Olympus BX61 upright microscope and images were recorded using a F-view II monochrome camera. The production of acetylated cellulosic polymer (ACP) was detected by the *in vivo* staining of colonies with Calcofluor (fluorescent brightener 28, Sigma) and this stain was visualised with fluorescence microscopy. Colonies of *P. fluorescens* were grown for 48 hours on KB plates containing Calcofluor, and several colonies were resuspended in 100 µL of distilled water. 10 µL of this solution was pipetted onto a microscope slide and used for microscopy. Stained cells were visualised using a 60x objective following DAPI excitation.

The localisation of fluorescently tagged proteins was visualised *in vivo* by fluorescent microscopy. Strains were then prepared for microscopy by inoculation of single colonies in 6 mL of sterile M9 media supplemented with 0.4% glycerol (with appropriate antibiotics to prevent loss of the mini-Tn7 vector) and grown overnight (160 rpm, 16 h). The culture was then subcultured in fresh M9-glycerol media (without antibiotics) resulting in an OD_{600} ~0.05 culture. This was then incubated with shaking for 60 minutes, after which 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce expression of the tagged genes. The culture was then returned to the incubator for 2 hours until an OD_{600} of approximately 0.3 was reached. In order to increase the density of cells for microscopy, a 1 mL aliquot of culture was centrifuged and resuspended in 100 µL of M9-glycerol media. Agarose pads were prepared consisting of M9-glycerol media with 1% w/v agarose. A 3 µL aliquot of resuspended culture of induced cells was pipetted onto the pads, which were left to dry and then covered with a coverslip. Cells were visualised using an Olympus BX61 upright microscope and both phase contrast and fluorescence images were recorded using a F-view II monochrome camera. Fluorescence images were observed using a constant 3500 ms exposure time across all strains. All images were processed using FIJI (Schindelin et al. 2012) to obtain measures of fluorescence and measures of co-localisation. Measures of fluorescence levels were obtained by calculating the mean gray values of cell transects observed under fluorescence

553 microscopy. Four randomly picked cells were analyzed across two independent images for
554 each biological replicate, with three biological replicates analyzed per genotype. Co-
555 localisation analyses were prepared by generating the mean van Steensel's distribution of
556 two fields of view for each biological replicate. At least three biological replicates were used
557 over two separate experiments in co-localisation analyses.
558
559
560
561
562
563

Acknowledgements

This work was supported in part by Marsden Fund Council from New Zealand Government funding (administered by the Royal Society of New Zealand) and the New Zealand Institute for Advanced Study. We thank Jenna Gallie and Peter Lind for discussion and comments on the manuscript and Heather Hendrickson for assistance with microscopy.

Competing interests

The authors declare no competing interests

573 **References**

574

575 Aldridge P, Paul R, Goymer P, Rainey P, Jenal U. 2003. Role of the GGDEF
576 regulator PleD in polar development of *Caulobacter crescentus*. *Mol*
577 *Microbiol* **47**(6): 1695-1708.
578 Amikam D, Galperin MY. 2006. PilZ domain is part of the bacterial c-di-GMP
579 binding protein. *Bioinformatics* **22**(1): 3-6. doi:
580 10.1093/bioinformatics/bti739
581 Andersen JB, Sternberg C, Poulsen LK, Bjorn SP, Givskov M, Molin S. 1998.
582 New unstable variants of green fluorescent protein for studies of
583 transient gene expression in bacteria. *Applied and Environmental*
584 *Microbiology* **64**(6): 2240-2246.
585 Annala MJ, Parker BC, Zhang W, Nykter M. 2013. Fusion genes and their
586 discovery using high throughput sequencing. *Cancer Letters* **340**(2):
587 192-200. doi: 10.1016/j.canlet.2013.01.011
588 Bailey SF, Bataillon T. 2016. Can the experimental evolution programme help
589 us elucidate the genetic basis of adaptation in nature? *Mol Ecol* **25**(1):
590 203-218.
591 Bantinaki E, Kassen R, Knight CG, Robinson Z, Spiers AJ, Rainey PB. 2007.
592 Adaptive divergence in experimental populations of *Pseudomonas*
593 *fluorescens*. III. Mutational origins of wrinkly spreader diversity.
594 *Genetics* **176**(1): 441-453. doi: 10.1534/genetics.106.069906
595 Bashton M, Chothia C. 2007. The generation of new protein functions by the
596 combination of domains. *Structure* **15**(1): 85-99. doi:
597 10.1016/j.str.2006.11.009

598 Bergthorsson U, Andersson DI, Roth JR. 2007. Ohno's dilemma: evolution of
599 new genes under continuous selection. *P Natl Acad Sci USA* **104**(43):
600 17004-17009. doi: 10.1073/pnas.0707158104
601 Bertani G. 1951. Studies on lysogenesis. I. The mode of phage liberation by
602 lysogenic *Escherichia coli*. *J Bacteriol* **62**(3): 293-300.
603 Blanka A, Duvel J, Dotsch A, Klinkert B, Abraham WR, Kaever V, Ritter C,
604 Narberhaus F, Haussler S. 2015. Constitutive production of c-di-GMP
605 is associated with mutations in a variant of *Pseudomonas aeruginosa*
606 with altered membrane composition. *Sci Signal* **8**(372). doi:
607 10.1126/scisignal.2005943
608 Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012. Genomic analysis of a
609 key innovation in an experimental *Escherichia coli* population. *Nature*
610 **489**(7417): 513-518. doi: 10.1038/nature11514
611 Bolte S, Cordelieres FP. 2006. A guided tour into subcellular colocalization
612 analysis in light microscopy. *Journal of Microscopy* **224**: 213-232. doi:
613 10.1111/j.1365-2818.2006.01706.x
614 Bridges C. 1936. The bar "gene" a duplication. *Science* **83**(2148): 210-211.
615 Brosius J. 1991. Retroposons--seeds of evolution. *Science* **251**(4995): 753-
616 753. doi: 10.1126/science.1990437
617 Buljan M, Frankish A, Bateman A. 2010. Quantifying the mechanisms of
618 domain gain in animal proteins. *Genome Biol* **11**(7): R74. doi:
619 10.1186/gb-2010-11-7-r74
620 Byun-McKay SA, Geeta R. 2007. Protein subcellular relocalization: a new
621 perspective on the origin of novel genes. *Trends in Ecology & Evolution*
622 **22**(7): 338-344. doi: 10.1016/j.tree.2007.05.002

623 Chan C, Paul R, Samoray D, Amiot NC, Giese B, Jenal U, Schirmer T. 2004.
624 Structural basis of activity and allosteric control of diguanylate cyclase.
625 *P Natl Acad Sci USA* **101**(49): 17084-17089. doi:
626 10.1073/pnas.0406134101
627 Chen SD, Krinsky BH, Long MY. 2013. New genes as drivers of phenotypic
628 evolution. *Nat Rev Genet* **14**(9): 645-660. doi: 10.1038/nrg3521
629 Choi K-H, Schweizer HP. 2006. mini-Tn7 insertion in bacteria with single
630 attTn7 sites: example *Pseudomonas aeruginosa*. *Nature Protocols*
631 **1**(1): 153-161. doi: 10.1038/nprot.2006.24
632 Chou SH, Galperin MY. 2016. Diversity of cyclic di-GMP-binding proteins and
633 mechanisms. *J Bacteriol* **198**(1): 32-46. doi: 10.1128/JB.00333-15
634 Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human
635 genome evolution. *Nat Rev Genet* **10**(10): 691-703. doi:
636 10.1038/nrg2640
637 Courseaux A, Nahon JL. 2001. Birth of two chimeric genes in the Hominidae
638 lineage. *Science* **291**(5507): 1293-1297. doi: 10.1126/science.1057284
639 Dai HZ, Chen Y, Chen SD, Mao QY, Kennedy D, Landback P, Eyre-Walker A,
640 Du W, Long MY. 2008. The evolution of courtship behaviors through
641 the origination of a new gene in Drosophila. *P Natl Acad Sci USA*
642 **105**(21): 7478-7483. doi: 10.1073/pnas.0800693105
643 De N, Pirruccello M, Krasteva PV, Bae N, Raghavan RV, Sondermann H.
644 2008. Phosphorylation-independent regulation of the diguanylate
645 cyclase WspR. *Plos Biol* **6**(3): 601-617. doi:
646 10.1371/journal.pbio.0060067

647 Ding Y, Zhou Q, Wang W. 2012. Origins of new genes and evolution of their
648 novel functions. *Annual Review of Ecology, Evolution, and Systematics*
649 **43**: 345-363. doi: 10.1146/annurev-ecolsys-110411-160513
650 Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of
651 eukaryotic genomes. *Annu Rev Genet* **41**: 331-368. doi:
652 10.1146/annurev.genet.40.110405.090448
653 Fu BD, Chen M, Zou M, Long MY, He SP. 2010. The rapid generation of
654 chimerical genes expanding protein diversity in zebrafish. *Bmc*
655 *Genomics* **11**. doi: 10.1186/1471-2164-11-657
656 Galperin MY. 2004. Bacterial signal transduction network in a genomic
657 perspective. *Environmental Microbiology* **6**(6): 552-567. doi:
658 10.1111/j.1462-2920.2004.00633.x
659 Goymer P, Kahn SG, Malone JG, Gehrig SM, Spiers AJ, Rainey PB. 2006.
660 Adaptive divergence in experimental populations of Pseudomonas
661 fluorescens. II. Role of the GGDEF regulator WspR in evolution and
662 development of the wrinkly spreader phenotype. *Genetics* **173**(2): 515-
663 526. doi: 10.1534/genetics.106.055863
664 Haldane J. 1932. The time of action of genes, and its bearing on some
665 evolutionary problems. *American Naturalist* **66**(702): 5-24.
666 Hammerschmidt K, Rose CJ, Kerr B, Rainey PB. 2014. Life cycles, fitness
667 decoupling and the evolution of multicellularity. *Nature* **515**(7525): 75-
668 79. doi: 10.1038/nature13884
669 Hanahan D. 1983. Studies on transformation of *Escherichia coli* with
670 plasmids. *J Mol Biol* **166**(4): 557-580.

671 Hengge R. 2009. Principles of c-di-GMP signalling in bacteria. *Nat Rev*
672 *Microbiol* **7**(4): 263-273. doi: 10.1038/nrmicro2109
673 Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989. Site-directed
674 mutagenesis by overlap extension using the polymerase chain
675 reaction. *Gene* **77**(1): 51-59.
676 Huangyutitham V, Guvener ZT, Harwood CS. 2013. Subcellular clustering of
677 the phosphorylated WspR response regulator protein stimulates its
678 diguanylate cyclase activity. *Mbio* **4**(3). doi: 10.1128/mBio.00242-13
679 Jenal U, Malone J. 2006. Mechanisms of cyclic-di-GMP signaling in bacteria.
680 *Annu Rev Genet* **40**: 385-407. doi:
681 10.1146/annurev.genet.40.110405.090423
682 Jin J, Xie XY, Chen C, Park JG, Stark C, James DA, Olhovsky M, Linding R,
683 Mao YY, Pawson T. 2009. Eukaryotic protein domains as functional
684 units of cellular evolution. *Sci Signal* **2**(98): ra76. doi:
685 10.1126/scisignal.2000546
686 Jones CD, Custer AW, Begun DJ. 2005. Origin and evolution of a chimeric
687 fusion gene in *Drosophila subobscura*, *D. madeirensis* and *D. guanche*.
688 *Genetics* **170**(1): 207-219. doi: 10.1534/genetics.104.037283
689 Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes.
690 *Genome Res* **20**(10): 1313-1326. doi: 10.1101/gr.101386.109
691 Katju V, Lynch M. 2003. The structure and early evolution of recently arisen
692 gene duplicates in the *Caenorhabditis elegans* genome. *Genetics*
693 **165**(4): 1793-1803.

694 Katju V, Lynch M. 2006. On the formation of novel genes by duplication in the
695 *Caenorhabditis elegans* genome. *Molecular Biology and Evolution*
696 **23**(5): 1056-1067. doi: 10.1093/molbev/msj114
697 King EO, Ward MK, Raney DE. 1954. Two simple media for the
698 demonstration of pyocyanin and fluorescin. *The Journal of laboratory*
699 *and clinical medicine* **44**(2): 301-307.
700 Kitten T, Kinscherf TG, McEvoy JL, Willis DK. 1998. A newly identified
701 regulator is required for virulence and toxin production in
702 Pseudomonas syringae. *Molecular Microbiology* **28**(5): 917-929. doi:
703 10.1046/j.1365-2958.1998.00842.x
704 Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting
705 transmembrane protein topology with a hidden Markov model:
706 Application to complete genomes. *J Mol Biol* **305**(3): 567-580. doi:
707 10.1006/jmbi.2000.4315
708 Lind PA, Farr AD, Rainey PB. 2015. Experimental evolution reveals hidden
709 diversity in evolutionary pathways. *Elife* **4**. doi: 10.7554/eLife.07074
710 Lind PA, Farr AD, Rainey PB. 2016. Evolutionary convergence in
711 experimental *Pseudomonas populations*. *ISME Journal* ***In press***.
712 Long M, Betran E, Thornton K, Wang W. 2003. The origin of new genes:
713 glimpses from the young and old. *Nat Rev Genet* **4**(11): 865-875. doi:
714 10.1038/nrg1204
715 Long M, Wang W, Zhang JM. 1999. Origin of new genes and source for N-
716 terminal domain of the chimerical gene, *jingwei*, in *Drosophila*. *Gene*
717 **238**(1): 135-141.

718 Long MY, Langley CH. 1993. Natural selection and the origin of *jingwei*, a
719 chimeric processed functional gene in *Drosophila*. *Science* **260**(5104):
720 91-95. doi: 10.1126/science.7682012
721 Long MY, Van Kuren NW, Chen SD, Vibranovski MD. 2013. New gene
722 evolution: little did we know. *Annual Review of Genetics, Vol 47* **47**:
723 307-333. doi: 10.1146/annurev-genet-111212-133301
724 Lynch M, Conery JS. 2000. The evolutionary fate and consequences of
725 duplicate genes. *Science* **290**(5494): 1151-1155.
726 Malone JG, Williams R, Christen M, Jenal U, Spiers AJ, Rainey PB. 2007.
727 The structure-function relationship of WspR, a Pseudomonas
728 fluorescens response regulator with a GGDEF output domain.
729 *Microbiology-SGM* **153**: 980-994. doi: 10.1099/mic.0.2006/002824-0
730 Mao XZ, Ma Q, Zhou C, Chen X, Zhang HY, Yang JC, Mao FL, Lai W, Xu Y.
731 2014. DOOR 2.0: presenting operons and their functions through
732 dynamic and integrated views. *Nucleic Acids Res* **42**(D1): D654-D659.
733 doi: 10.1093/nar/gkt1048
734 Marques AC, Vinckenbosh N, Brawand D, Kaessmann H. 2008. Functional
735 diversification of duplicate genes through subcellular adaptation of
736 encoded proteins. *Genome Biol* **9**(3). doi: 10.1186/gb-2008-9-3-r54
737 Marsh JA, Teichmann SA. 2010. How do proteins gain new domains?
738 *Genome Biol* **11**(7): 126. doi: 10.1186/gb-2010-11-7-126
739 Massie JP, Reynolds EL, Koestler BJ, Cong JP, Agostoni M, Waters CM.
740 2012. Quantification of high-specificity cyclic diguanylate signaling. *P*
741 *Natl Acad Sci USA* **109**(31): 12746-12751. doi:
742 10.1073/pnas.1115663109

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**(3): 1041-1053. doi: 10.1534/genetics.109.107110

Mitra A, Kesarwani AK, Pal D, Nagaraja V. 2011. WebGeSTer DB-a transcription terminator database. *Nucleic Acids Res* **39**: D129-D135. doi: 10.1093/nar/gkq971

Moran JV, DeBerardinis RJ, Kazazian HH. 1999. Exon shuffling by L1 retrotransposition. *Science* **283**(5407): 1530-1534.

Nasvall J, Sun L, Roth JR, Andersson DI. 2012. Real-time evolution of new genes by innovation, amplification, and divergence. *Science* **338**(6105): 384-387. doi: 10.1126/science.1226521

O'Connor JR, Kuwada NJ, Huangyutitham V, Wiggins PA, Harwood CS. 2012. Surface sensing and lateral subcellular localization of WspA, the receptor in a chemosensory-like system leading to c-di-GMP production. *Mol Microbiol* **86**(3): 720-729. doi: 10.1111/mmi.12013

Ohno S. 1970. *Evolution by Gene Duplication*. Springer, New York.

Pasek S, Risler JL, Brezellec P. 2006. Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* **22**(12): 1418-1423. doi: 10.1093/bioinformatics/btl135

Patthy L. 2003. Modular assembly of genes and the evolution of new functions. *Genetica* **118**(2-3): 217-231. doi: 10.1023/a:1024182432483

Paul R, Abel S, Wassmann P, Beck A, Heerklotz H, Jenal U. 2007. Activation of the diguanylate cyclase PleD by phosphorylation-mediated

768 dimerization. *J Biol Chem* **282**(40): 29170-29177. doi:
769 10.1074/jbc.M704702200
770 Peisajovich SG, Garbarino JE, Wei P, Lim WA. 2010. Rapid diversification of
771 cell signaling phenotypes by modular domain recombination. *Science*
772 **328**(5976): 368-372. doi: 10.1126/science.1182376
773 Rainey PB. 1999. Adaptation of *Pseudomonas fluorescens* to the plant
774 rhizosphere. *Environmental Microbiology* **1**(3): 243-257. doi:
775 10.1046/j.1462-2920.1999.00040.x
776 Rainey PB, Travisano M. 1998. Adaptive radiation in a heterogeneous
777 environment. *Nature* **394**(6688): 69-72. doi: 10.1038/27900
778 Ranz JM, Parsch J. 2012. Newly evolved genes: Moving from comparative
779 genomics to functional studies in model systems. *Bioessays* **34**(6):
780 477-483. doi: 10.1002/bies.201100177
781 Ranz JM, Ponce AR, Hartl DL, Nurminsky D. 2003. Origin and evolution of a
782 new gene expressed in the *Drosophila* sperm axoneme. *Genetica*
783 **118**(2-3): 233-244.
784 Rogers RL, Bedford T, Hartl DL. 2009. Formation and longevity of chimeric
785 and duplicate genes in *Drosophila melanogaster*. *Genetics* **181**(1):
786 313-322. doi: 10.1534/genetics.108.091538
787 Rogers RL, Bedford T, Lyons AM, Hartl DL. 2010. Adaptive impact of the
788 chimeric gene *Quetzalcoatl* in *Drosophila melanogaster*. *P Natl Acad*
789 *Sci USA* **107**(24): 10943-10948. doi: 10.1073/pnas.1006503107
790 Rogers RL, Hartl DL. 2012. Chimeric genes as a source of rapid evolution in
791 *Drosophila melanogaster*. *Molecular Biology and Evolution* **29**(2): 517-
792 529. doi: 10.1093/molbev/msr184

793 Romling U, Galperin MY, Gomelsky M. 2013. Cyclic di-GMP: the first 25 years
794 of a universal bacterial second messenger. *Microbiology and Molecular*
795 *Biology Reviews* **77**(1): 1-52. doi: 10.1128/mmbr.00043-12
796 Ross P, Weinhouse H, Aloni Y, Michaeli D, Weinberger-Ohana P, Mayer R,
797 Braun S, de Vroom E, van der Marel GA, van Boom JH et al. 1987.
798 Regulation of cellulose synthesis in Acetobacter xylinum by cyclic
799 diguanylic acid. *Nature* **325**(6101): 279-281.
800 Sambrook J, Russell DW. 2001. *Molecular cloning. A laboratory manual.* Cold
801 Spring Harbour Laboratory Press Cold Spring Harbour, New York.
802 Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T,
803 Preibisch S, Rueden C, Saalfeld S, Schmid B et al. 2012. Fiji: an open-
804 source platform for biological-image analysis. *Nature Methods* **9**(7):
805 676-682. doi: 10.1038/nmeth.2019
806 Schirmer T, Jenal U. 2009. Structural and mechanistic determinants of c-di-
807 GMP signalling. *Nat Rev Microbiol* **7**(10): 724-735. doi:
808 10.1038/nrmicro2203
809 Silby MW, Cerdeno-Tarraga AM, Vernikos GS, Giddens SR, Jackson RW,
810 Preston GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SAC et al.
811 2009. Genomic and genetic analyses of diversity and plant interactions
812 of Pseudomonas fluorescens. *Genome Biol* **10**(5). doi: 10.1186/gb-
813 2009-10-5-r51
814 Singh D, Chan JM, Zoppoli P, Niola F, Sullivan R, Castano A, Liu EM, Reichel
815 J, Porrati P, Pellegatta S et al. 2012. Transforming fusions of FGFR
816 and TACC genes in human glioblastoma. *Science* **337**(6099): 1231-
817 1235. doi: 10.1126/science.1220834

818 Spiers AJ, Bohannon J, Gehrig SM, Rainey PB. 2003. Biofilm formation at the
819 air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly
820 spreader requires an acetylated form of cellulose. *Mol Microbiol* **50**(1):
821 15-27. doi: 10.1046/j.1365-2958.2003.03670.x
822 Spiers AJ, Kahn SG, Bohannon J, Travisano M, Rainey PB. 2002. Adaptive
823 divergence in experimental populations of *Pseudomonas fluorescens*.
824 I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics*
825 **161**(1): 33-46.
826 Sturtevant A. 1925. The effects of unequal crossing over at the bar locus in
827 drosophila. *Genetics* **10**(117): 117-147.
828 Tal R, Wong HC, Calhoon R, Gelfand D, Fear AL, Volman G, Mayer R, Ross
829 P, Amikam D, Weinhouse H et al. 1998. Three *cdg* operons control
830 cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: Genetic
831 organization and occurrence of conserved domains in isoenzymes. *J*
832 *Bacteriol* **180**(17): 4416-4425.
833 Tautz D, Domazet-Loso T. 2011. The evolutionary origin of orphan genes.
834 *Nature Reviews Genetics* **12**(10): 692-702. doi: 10.1038/nrg3053
835 Thomson TM, Lozano JJ, Loukili N, Carrio R, Serras F, Cormand B, Valeri M,
836 Diaz VM, Abril J, Burset M et al. 2000. Fusion of the human gene for
837 the polyubiquitination coeffector UEV1 with *Kua*, a newly identified
838 gene. *Genome Res* **10**(11): 1743-1756. doi: 10.1101/gr.GR-1405R
839 van Steensel B, van Binnendijk EP, Hornsby CD, van der Voort HTM,
840 Krozowski ZS, de Kloet ER, van Driel R. 1996. Partial colocalization of
841 glucocorticoid and mineralocorticoid receptors in discrete

842 compartments in nuclei of rat hippocampus neurons. *Journal of Cell*
843 *Science* **109**: 787-792.
844 Volff JN. 2006. Turning junk into gold: domestication of transposable elements
845 and the creation of new genes in eukaryotes. *Bioessays* **28**(9): 913-
846 922. doi: 10.1002/bies.20452
847 Wang W, Brunet FG, Nevo E, Long M. 2002. Origin of *sphinx*, a young
848 chimeric RNA gene in *Drosophila melanogaster*. *P Natl Acad Sci USA*
849 **99**(7): 4448-4453. doi: 10.1073/pnas.072066399
850 Wang W, Zhang JM, Alvarez C, Llopart A, Long M. 2000. The origin of the
851 Jingwei gene and the complex modular structure of its parental gene,
852 yellow emperor, in Drosophila melanogaster. *Molecular Biology and*
853 *Evolution* **17**(9): 1294-1301.
854 Wang W, Zheng HK, Fan CZ, Li J, Shi JJ, Cai ZQ, Zhang GJ, Liu DY, Zhang
855 JG, Vang S et al. 2006. High rate of chimeric gene origination by
856 retroposition in plant genomes. *Plant Cell* **18**(8): 1791-1802. doi:
857 10.1105/tpc.106.041905
858 Wang XJ, Huang Y, Lavrov DV, Gu X. 2009. Comparative study of human
859 mitochondrial proteome reveals extensive protein subcellular
860 relocalization after gene duplications. *Bmc Evolutionary Biology* **9**. doi:
861 10.1186/1471-2148-9-275
862 Wassmann P, Chan C, Paul R, Beck A, Heerklotz H, Jenal U, Schirmer T.
863 2007. Structure of BeF3--modified response regulator PleD:
864 Implications for diguanylate cyclase activation, catalysis, and feedback
865 inhibition. *Structure* **15**(8): 915-927. doi: 10.1016/j.str.2007.06.016

866 Whiteside MD, Winsor GL, Laird MR, Brinkman FSL. 2013. OrthologeDB: a
867 bacterial and archaeal orthology resource for improved comparative
868 genomic analysis. *Nucleic Acids Res* **41**(D1): D366-D376. doi:
869 10.1093/nar/gks1241
870 Winsor GL, Lam DKW, Fleming L, Lo R, Whiteside MD, Yu NY, Hancock
871 REW, Brinkman FSL. 2011. Pseudomonas Genome Database:
872 improved comparative analysis and population genomics capability for
873 *Pseudomonas* genomes. *Nucleic Acids Res* **39**: D596-D600. doi:
874 10.1093/nar/gkq869
875 Yeh SD, Do T, Chan C, Cordova A, Carranza F, Yamamoto EA, Abbassi M,
876 Gandasetiawan KA, Librado P, Damia E et al. 2012. Functional
877 evidence that a recently evolved *Drosophila* sperm-specific gene
878 boosts sperm competition. *P Natl Acad Sci USA* **109**(6): 2043-2048.
879 doi: 10.1073/pnas.1121327109
880 Zhang JM, Dean AM, Brunet F, Long MY. 2004. Evolving protein functional
881 diversity in new genes of *Drosophila*. *P Natl Acad Sci USA* **101**(46):
882 16246-16250. doi: 10.1073/pnas.0407066101
883 Zhang JM, Yang HY, Long MY, Li LM, Dean AM. 2010. Evolution of
884 enzymatic activities of testis-specific short-chain
885 dehydrogenase/reductase in *Drosophila*. *Journal of Molecular*
886 *Evolution* **71**(4): 241-249. doi: 10.1007/s00239-010-9384-5
887 Zhang Y, Lu S, Zhao S, Zheng X, Long M, Wei L. 2009. Positive selection for
888 the male functionality of a co-retroposed gene in the hominoids. *Bmc*
889 *Evolutionary Biology* **9**. doi: 10.1186/1471-2148-9-252

890 Zhang YE, Long M. 2014. New genes contribute to genetic and phenotypic
891 novelties in human evolution. *Current Opinion in Genetics &*
892 *Development* **29**: 90-96. doi: 10.1016/j.gde.2014.08.013
893 Zhang YM, Rock CO. 2008. Membrane lipid homeostasis in bacteria. *Nat Rev*
894 *Microbiol* **6**(3): 222-233. doi: 10.1038/nrmicro1839
895 Zhao L, Saelao P, Jones CD, Begun DJ. 2014. Origin and spread of de novo
896 genes in *Drosophila melanogaster* populations. *Science* **343**(6172):
897 769-772. doi: 10.1126/science.1248286
898 Zhu K, Choi KH, Schweizer HP, Rock CO, Zhang YM. 2006. Two aerobic
899 pathways for the formation of unsaturated fatty acids in *Pseudomonas*
900 *aeruginosa*. *Mol Microbiol* **60**(2): 260-273. doi: 10.1111/j.1365-
901 2958.2006.05088.x
902
903
904
905
906
907
908
909
910
911
912

Figure Supplements

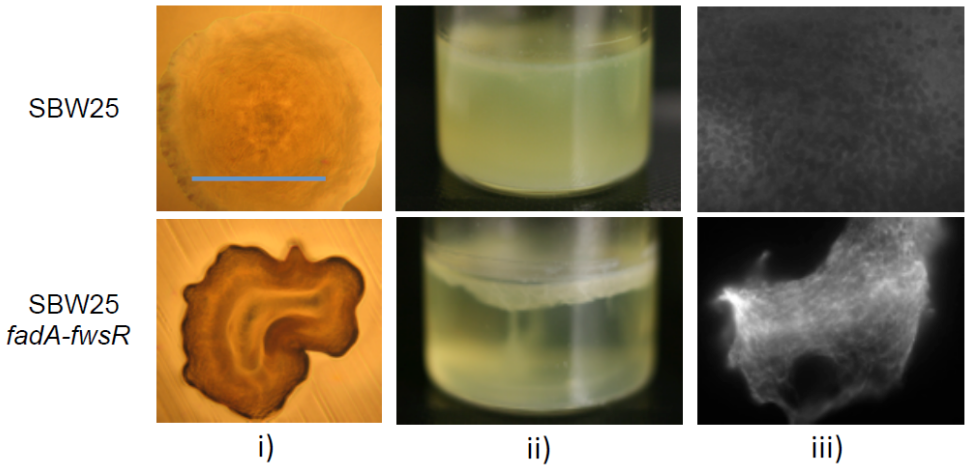
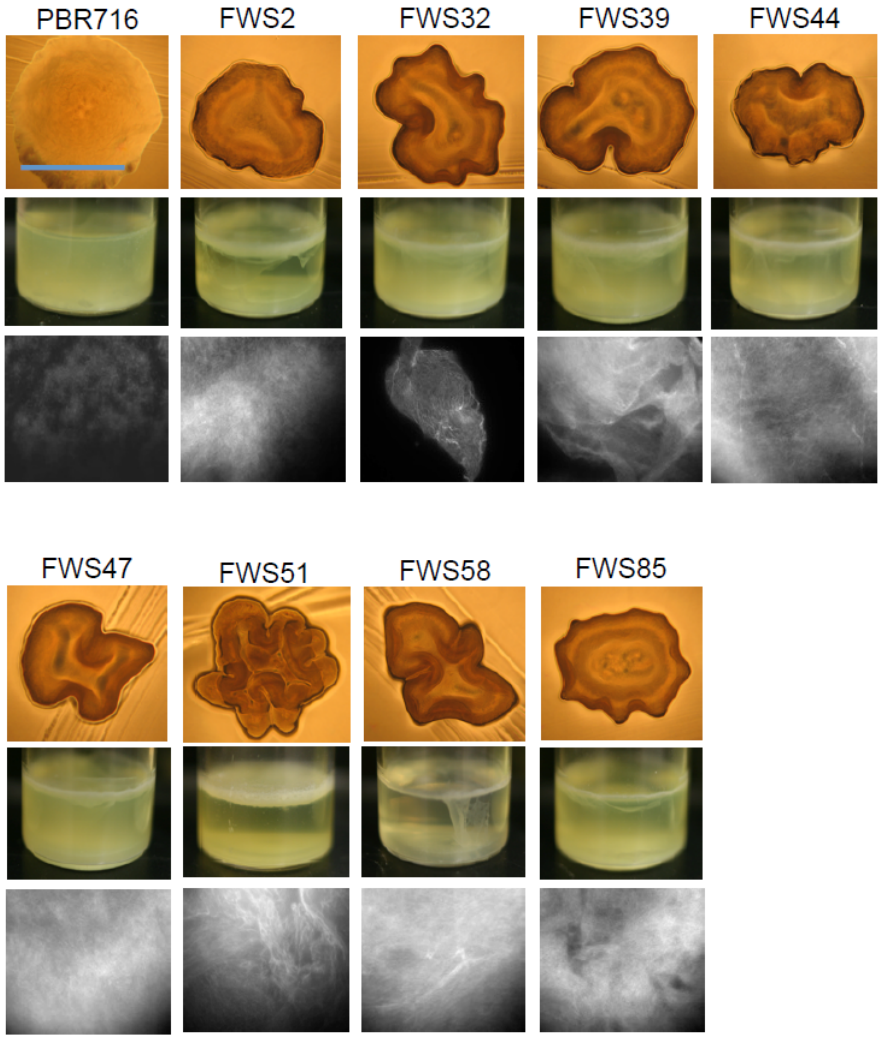


Figure 1 – figure supplement 1: The reconstruction of the *fadA-fwsR* mutation in ancestral backgrounds causes WS. Genotypes featuring *fadA-fwsR* (either naturally evolved or reconstructed in SM backgrounds) are characteristically WS in terms of i) morphology (visualised by light microscopy (10x objective) of ~32 h colonies grown on King's Medium B agar (KB), blue bar is ~2mm), ii) formation of mats at the air-liquid interface of microcosms (incubated statically at 28^{o}C for 3 days) and iii) calcofluor binding as indicative of ACP biosynthesis (visualised by fluorescent microscopy, 60x objective). SBW25 is ancestral *P. fluorescens* SBW25, SBW25 *fadA-fwsR* has the causal *fadA-fwsR* mutation reconstructed into the ancestral SM background.



924

925 **Figure 1 - figure supplement 2: The spectrum of *fadA-fwsR* fusions arising from SBW25 Δ *wsp***
926 ***aws mws* (PBR716) express the WS phenotype.** Numbered are the different isolates with *fadA-fwsR*
927 fusion mutations. The phenotypes of the *fadA-fwsR* fusions are characteristic WS in morphology,
928 colonisation of the air-liquid interface and ACP production (see Figure 1 – figure supplement 1 for
929 details of each image).

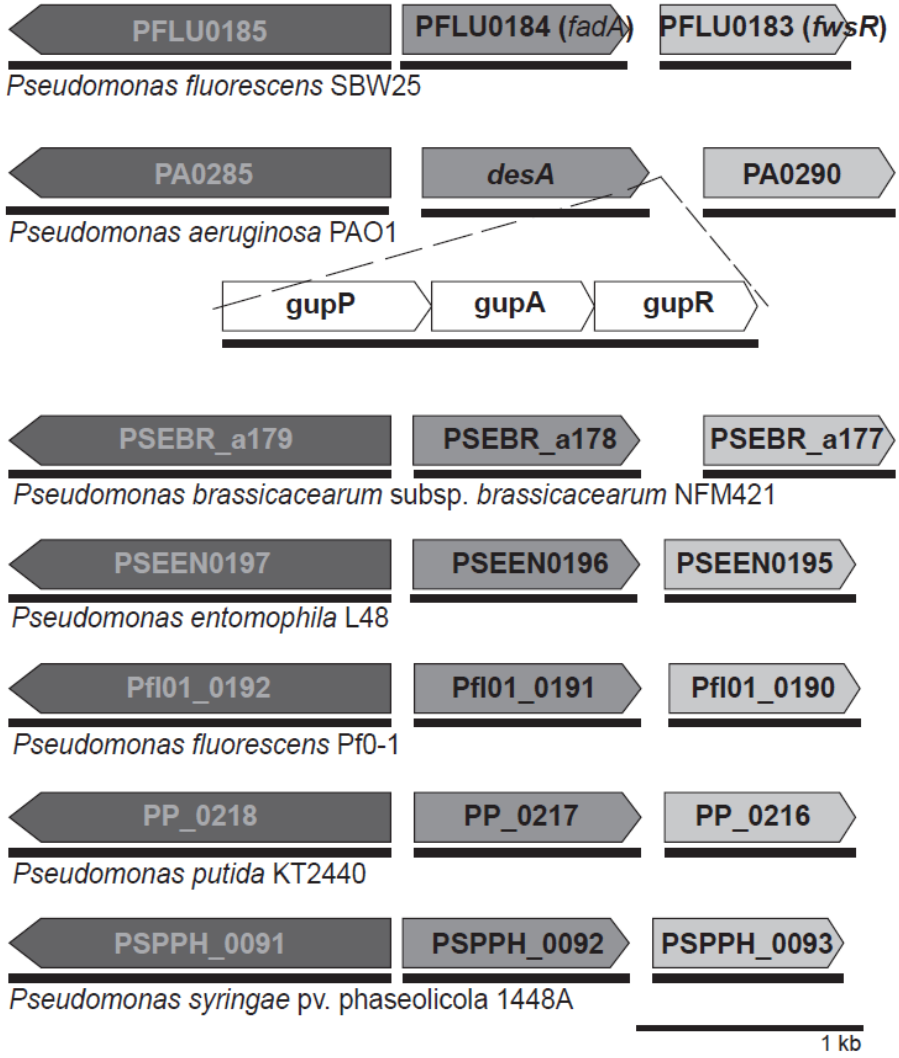


Figure 1 - Figure Supplement 3: Organisation of homologs of *fadA* and *fwsR* across *Pseudomonas* spp. The arrangement of neighbouring genes of the *fwsR* locus is conserved across many species of the genus *Pseudomonas*. Notable exceptions are strains of *P. aeruginosa*, in which the paralogs *fadA* and *fwsR* are separated by a single predicted operon of genes *gupPAR*. There is no evidence of recombination between *fadA* and *fwsR* in other strains. Genes sharing the same shade represent likely paralogs as determined by the *Ortholuge* web program (Whiteside et al. 2013), with thin black bars (underneath depicted ORFs) representing operons as predicted by the *DOOR* web program (Mao et al. 2014)

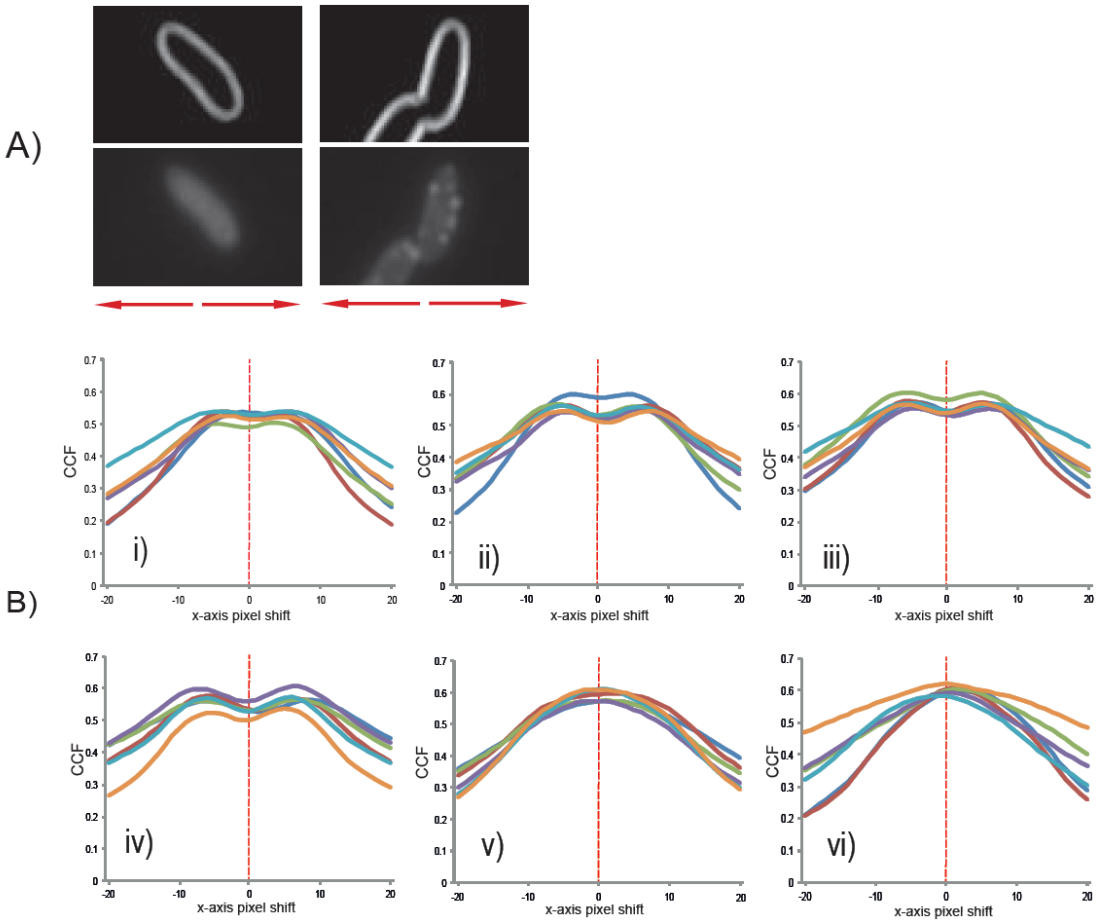


Figure 7 – figure supplement 1: van Steensel's distribution of colocalisation of fluorescent signal and the cell edge. A) The CCF represents the degree of colocalisation of the reference image (the marked edges of cells observed by phase contrast microscopy) and the fluorescent distribution of images. The edges of cells were marked using FIJI (Schindelin et al. 2012). Depicted are representative images of cells expressing *fwsR-gfp* (left) and *fadA-fwsR-gfp* (right). To generate van Steensel's distributions, the images are moved horizontally relative to each other and the CCF value is calculated using FIJI with the JACoP plugin (Bolte and Cordelieres 2006). **B)** Graphs represent van Steensel's distributions for i) *gfp*(-) control; ii) *gfp*(+) control; iii) *fwsR-gfp*; iv) *fadA-3X-fwsR-gfp*; v) *fadA-gfp*; and vi) *fadA-fwsR-gfp*. Graphs i, ii, iii and iv depict bimodal distributions indicative of a fluorescent distribution in between the reference markings of the cell edge. Distributions v and vi are unimodal and the distribution peaks approximately at a pixel-shift of 0, indicative of direct correlation of the fluorescent signal with the cell edge. Analyses of six biological replicates were performed per genotype, with each replicate the analysis of approximately 100 cells analysed per biological replicate.