



# *Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system*

Article

Accepted Version

Last edited version for re-submission

Taylor, T. B., Mulley, G., Dills, A. H., Alsohim, A. S., McGuffin, L. J., Studholme, D. J., Silby, M. W., Brockhurst, M. A., Johnson, L. J. and Jackson, R. W. (2015) Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system. *Science*, 347 (6225). pp. 1014-1017. ISSN 0036-8075 doi: <https://doi.org/10.1126/science.1259145>  
Available at <http://centaur.reading.ac.uk/39447/>

It is advisable to refer to the publisher's version if you intend to cite from the work.

Published version at: <http://dx.doi.org/10.1126/science.1259145>

To link to this article DOI: <http://dx.doi.org/10.1126/science.1259145>

Publisher: AAAS

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

**Title: Evolutionary resurrection of flagellar motility via rewiring of the  
nitrogen regulation system**

**Authors: Tiffany B. Taylor<sup>1†</sup>, Geraldine Mulley<sup>1†</sup>, Alexander H. Dills<sup>2</sup>, Abdullah S. Alsohim<sup>1,3</sup>, Liam J. McGuffin<sup>1</sup>, David J. Studholme<sup>4</sup>, Mark W. Silby<sup>2</sup>, Michael A. Brockhurst<sup>5</sup>, Louise J. Johnson<sup>1,\*</sup>, Robert W. Jackson<sup>1,6</sup>**

**Affiliations:**

<sup>1</sup>School of Biological Sciences, University of Reading, Whiteknights, Reading, RG6 6AJ, UK

<sup>2</sup>Department of Biology, University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, MA 02747, USA

<sup>3</sup>Department of Plant Production and Protection, Qassim University, Qassim, Saudi Arabia P.O. Box 6622

<sup>4</sup>College of Life and Environmental Sciences, University of Exeter, Stocker Road, Exeter, EX4 4QD, UK

<sup>5</sup>Department of Biology, University of York, Wentworth Way, York, YO10 5DD, UK

<sup>6</sup>The University of Akureyri, Borgir vid Nordurslod, IS-600 Akureyri, Iceland

\*Corresponding author: [L.J.Johnson@reading.ac.uk](mailto:L.J.Johnson@reading.ac.uk)

† These authors contributed equally to this work

1 **One Sentence Summary:**

2 Rapid repeatable rewiring of regulatory networks: a nitrogen regulatory gene evolves a  
3 new function, restoring flagella to immotile bacteria.

4  
5 **Abstract:**

6 A central process in evolution is the recruitment of genes to regulatory networks. We  
7 engineered immotile strains of the bacterium *Pseudomonas fluorescens* that lack  
8 flagella due to deletion of the regulatory gene *fleQ*. Under strong selection for motility,  
9 these bacteria consistently regained flagella within 96 hours via a two-step evolutionary  
10 pathway. Step 1 mutations increase intracellular levels of phosphorylated NtrC, a distant  
11 homologue of FleQ, which begins to commandeer control of the *fleQ* regulon at the cost  
12 of disrupting nitrogen uptake and assimilation. Step 2 is a switch-of-function mutation  
13 that redirects NtrC away from nitrogen uptake and towards its novel function as a  
14 flagellar regulator. Our results demonstrate that natural selection can rapidly rewire  
15 regulatory networks in very few, repeatable mutational steps.

16

17 **Main Text:**

18 A longstanding evolutionary question concerns how the duplication and  
19 recruitment of genes to regulatory networks facilitates their expansion (1), and how  
20 networks gain mutational robustness and evolvability (2). Bacteria respond to diverse  
21 environments using a vast range of specialised regulatory pathways, predominantly  
22 two-component systems (3), which are the result of adaptive radiations within gene

23 families. Due to past cycles of gene duplication, divergence and horizontal genetic  
24 transfer, there is often extensive homology between the components of different  
25 pathways (4), raising the possibility of cross-talk or redundancy between pathways (5).  
26 Here we monitor the recovery of microbial populations from a catastrophic gene  
27 deletion: bacteria engineered to lack a particular function are exposed to environments  
28 that impose strong selection to re-evolve it, sometimes by recruitment of new genes to  
29 regulatory networks (6, 7, 8, 9).

30 In the plant-associated soil bacterium *P. fluorescens*, the master regulator of  
31 flagellar synthesis is FleQ (also called AdnA), a  $\sigma^{54}$ -dependent enhancer binding protein  
32 (EBP) that activates transcription of genes required for flagellum biosynthesis (10, 11).  
33 The starting *P. fluorescens* strain is AR2; this strain lacks flagella, due to deletion of  
34 *fleQ*, and is unable to move by spreading motility due to mutation of viscosin synthase  
35 (*viscB*), resulting in a distinctive, point-like colony morphology on spreading motility  
36 medium (SMM) (12) (Figure 1A). We grew replicate populations of AR2 on SMM; when  
37 local nutrients became depleted, starvation imposed strong selection to re-evolve  
38 motility. To demonstrate that this finding was not strain-specific, these experiments were  
39 replicated in a different strain of *P. fluorescens*, Pf0-2x. This strain is a  $\Delta fleQ$  variant of  
40 Pf0-1, already viscosin-deficient, and is thus unable to move by spreading or swimming  
41 motility.

42 After 96 hours incubation of AR2 and Pf0-2x at room temperature on SMM, two  
43 breakout mutations were visible conferring first slow (AR2S and Pf0-2xS) and then fast  
44 (AR2F and Pf0-2xF) spreading over the agar surface (Fig. 1A). The AR2F strain  
45 produces flagella, but we could not detect flagella in EM samples for AR2S (Fig. 1B).

46 Genome resequencing revealed a single nucleotide point mutation in *ntrB* in strain  
47 AR2S, causing an amino acid substitution within the PAS domain of the histidine kinase  
48 (HK) sensor NtrB (T97P). The fast-spreading strain AR2F had acquired an additional  
49 point mutation in the  $\sigma^{54}$ -dependent EBP gene *ntrC*, which alters an amino acid  
50 (R442C) within the DNA-binding domain (Table 1 & S2).

51 NtrB and NtrC comprise a two-component system: under nitrogen limitation NtrB  
52 phosphorylates NtrC, which activates transcription of genes required for nitrogen uptake  
53 and metabolism. To determine how mutations in this separate regulatory pathway  
54 restored motility in the absence of FleQ, we performed microarray and qRT-PCR  
55 analyses of the ancestral and evolved strains (Fig. S1 & Table S1). The expression of  
56 genes required for flagellum biosynthesis and chemotaxis was abolished in AR2  
57 compared to wild-type SBW25 (Fig. 2A). The *ntrB* mutation in AR2S partially restores  
58 the expression of flagellar genes, and over-activates the expression of genes involved  
59 in nitrogen regulation, uptake and metabolism. The subsequent *ntrC* mutation in AR2F  
60 reduces the expression of nitrogen uptake and metabolism genes, while further up-  
61 regulating flagellar and chemotaxis gene expression to wild-type levels (Fig. 2B). While  
62 AR2S and AR2F showed higher growth rates than the ancestor in LB medium (the  
63 medium on which the mutants arose; Tukey-Kramer HSD test, growth in LB compared  
64 to AR2: AR2S,  $P < 0.001$ ; AR2F,  $P < 0.001$ ) (Fig. 1C), both mutants grew poorly in  
65 minimal medium with ammonium as the sole nitrogen source (Tukey-Kramer HSD test,  
66 growth in M9 + ammonium compared to AR2: AR2S,  $P < 0.001$ ; AR2F,  $P = 0.001$ ). This  
67 is likely to be the result of ammonium toxicity due to the strong up-regulation of genes

68 involved in ammonium uptake and assimilation, indicating a pleiotropic cost of this  
69 adaptation.

70 Sequencing of the *ntrBC* locus from independently evolved replicate strains  
71 showed that evolution often followed parallel trajectories in both AR2 and Pf0-2x:  
72 mutation of *ntrB* gave a slow-spreading strain and this was followed by mutation of *ntrC*  
73 yielding a fast-spreading strain (Table 1; Fig. 3A & C). While all Pf0-2xF replicates  
74 carried mutations in *ntrC*, several Pf0-2xS strains were not mutated in *ntrB*, suggesting  
75 an alternative evolutionary pathway. Genome resequencing of these strains revealed  
76 mutations in *glnA* or *glnK* (Table 1; Fig. 3B) likely to result in loss of function leading to  
77 abnormally high levels of phosphorylated NtrC: reduced ammonium assimilation by  
78 glutamine synthetase (*glnA*) would impose severe nitrogen limitation in the cell  
79 irrespective of nitrogen availability, whereas GlnK is a PII-protein that regulates both  
80 NtrB and glutamine synthetase activities.

81 These data suggest a predictable two-step evolutionary process: **Step 1**  
82 increases levels of phosphorylated NtrC, via either (a) a direct regulatory route with  
83 mutations in NtrB or GlnK, or (b) a physiological route with loss-of-function mutations  
84 reducing glutamine synthetase activity and causing NtrB activation, partially or  
85 intermittently reactivating the flagellar cascade. In **Step 2**, NtrC adapts to enhance  
86 activation of the flagellar genes and in doing so, becomes a less potent activator of  
87 nitrogen uptake genes. This model explains the microarray data and is consistent with  
88 the predicted structural effects of the mutations (Figs. 2 and 3). Specifically, for NtrB the  
89 structural changes are likely to either increase kinase or reduce phosphatase activity. In

90 support of this, the mutated NtrB(D228A) repeatedly emerging in Pf0-2xS resembles  
91 NtrB(D227A) in *P. aeruginosa* which constitutively activates the Ntr system (13).

92 NtrC is a distant homologue of FleQ, sharing 30% amino acid identity and the  
93 same three structural domains (TM-score > 0.7;  $P < 0.001$ ; Fig. 3D) (14): an N-terminal  
94 receiver domain, a conserved central  $\sigma^{54}$ -interacting domain, and a C-terminal DNA-  
95 binding domain containing a helix-turn-helix (HTH) motif flanked by highly disordered  
96 regions. We posit that an overabundance of phosphorylated NtrC activates transcription  
97 of flagellar genes through functional promiscuity (15). Consistent with this, the *ntrC*  
98 mutations in fast-spreading strains are predominantly located within or adjacent to the  
99 HTH domain and likely influence enhancer-binding specificity; one is a frameshift  
100 abolishing the HTH altogether (Fig. 3C). The evolved NtrC' must be constitutively  
101 phosphorylated by overactive NtrB to enable its new multipurpose role, with the result  
102 that flagellum biosynthesis and nitrogen regulation are probably no longer responsive to  
103 environmental stimuli. Consequently, there is a trade-off between nitrogen utilization  
104 and motility (Fig. 1A & C).

105 The flagellar regulatory network may have an unusually dynamic evolutionary  
106 history. Flagella are expensive to make, and not always advantageous. Pathogens  
107 expressing flagella can trigger strong immune responses in the host, so rapid transitions  
108 are seen over short timescales between unflagellate, multiflagellate and aflagellate  
109 states (16). This volatility is reflected in the structure of regulatory networks: in close  
110 relatives of *Pseudomonas*, *fleQ* appears not to be involved in flagellar gene expression  
111 (17), and in *Helicobacter pylori*, a gene of unknown function can be used as a “spare  
112 part” to permit motility in *flhB* mutants (18). Our results illustrate that trans-acting



113 mutations can contribute to gene network evolution (19), but that as predicted, such  
114 mutations bear severe pleiotropic costs (20, 21). Genes can retain the potential to take  
115 on the functions of long-diverged homologues, suggesting that some degree of  
116 evolutionary resilience is a consequence of regulatory pathways that evolve via gene  
117 duplication. While *de novo* origination of new functions in nature is likely to take longer  
118 and involve more mutational steps, this system enables us to understand the adaptive  
119 process in detail at the genetic and phenotypic level. Here we identified a novel,  
120 tractable model for gene network evolution and observed, in real time, the rewiring of  
121 gene networks to enable the incorporation of a modified component (NtrC') creating a  
122 novel regulatory function, by a highly repeatable two-step evolutionary pathway with the  
123 same point mutations often recurring in independent lineages.

124

## 125 **References:**

- 126 1. J. R. True, S. B. Carroll, Gene co-option in physiological and morphological  
127 evolution. *Annu. Rev. Cell. Dev. Bi.* **18**, 53-80 (2002).
- 128 2. M. Aldana, E. Balleza, S. Kauffman, O. Resendiz, Robustness and evolvability in  
129 genetic regulatory networks. *J. Theor. Biol.* **245**, 433-448 (2007).
- 130 3. J. A. Hoch, Two-component and phosphorelay signal transduction. *Curr. Opin.*  
131 *Microbiol.* **3**, 165-170 (2000).
- 132 4. S. A. Teichmann, M. M. Babu, Gene regulatory network growth by duplication.  
133 *Nat. Genet.* **36**, 492-496 (2004).

- 134 5. E. J. Capra, M. T. Laub, Evolution of two-component signal transduction  
135 systems. *Annu. Rev. Microbiol.* **66**, 325-347 (2012).
- 136 6. Z. D. Blount, C. Z. Borland, R. E. Lenski, Historical contingency and the evolution  
137 of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl.*  
138 *Acad. Sci. USA* **105**, 7899-7906 (2008).
- 139 7. J. R. Meyer, *et al.*, Repeatability and contingency in the evolution of a key  
140 innovation in phage lambda. *Science* **335**, 428-432 (2012).
- 141 8. B. Hall, The EBG system of *E. coli*: Origin and evolution of a novel  $\beta$ -  
142 galactosidase for the metabolism of lactose. *Genetica* **118**, 143-156 (2003).
- 143 9. D. Blank, L. Wolf, M. Ackermann, O. K. Silander, The predictability of molecular  
144 evolution during functional innovation. *Proc. Natl. Acad. Sci. USA* **111**, 3044-  
145 3049 (2014).
- 146 10. E. A. Robleto, I. López-Hernández, M. W. Silby, S. B. Levy, Genetic analysis of  
147 the AdnA regulon in *Pseudomonas fluorescens*: Nonessential role of flagella in  
148 adhesion to sand and biofilm formation. *J. Bacteriol.* **185**, 453-460 (2003).
- 149 11. N. Dasgupta, *et al.*, A four-tiered transcriptional regulatory circuit controls  
150 flagellar biogenesis in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **50**, 809-824  
151 (2003).
- 152 12. A. S. Alsohim, *et al.*, The biosurfactant viscosin produced by *Pseudomonas*  
153 *fluorescens* SBW25 aids spreading motility and plant growth promotion. *Environ.*  
154 *Microbiol.* **16**, 2267-2281 (2014).

- 155 13. W. Li, C.-D. Lu, Regulation of carbon and nitrogen utilization by CbrAB and  
156 NtrBC two-component systems in *Pseudomonas aeruginosa*. *J. Bacteriol.* **189**,  
157 5413-5420 (2007).
- 158 14. D. J. Studholme, R. Dixon, Domain architectures of  $\sigma^{54}$ -dependent transcriptional  
159 activators. *J. Bacteriol.* **185**, 1757-1767 (2003).
- 160 15. R. Wasseem, E. M. de Souza, M. G. Yates, F. D. Pedrosam, M. Buck, Two roles  
161 for integration host factor at an enhancer-dependent *nifA* promoter, *Mol.*  
162 *Microbiol.* **35**, 756-764 (2000).
- 163 16. E. Amiel, R. R. Lovewell, G. A. O'Toole, D. A. Hogan, B. Berwin, *Pseudomonas*  
164 *aeruginosa* evasion of phagocytosis is mediated by loss of swimming motility and  
165 is independent of flagellum expression. *Infect. Immun.* **78**, 2937-2945 (2010).
- 166 17. R. León, G. Espín, *flhDC*, but not *fleQ*, regulates flagella biogenesis in  
167 *Azotobacter vinelandii*, and is under AlgU and CydR negative control.  
168 *Microbiology* **154**, 1719-1728 (2008).
- 169 18. M. E. Wand, *et al.*, *Helicobacter pylori* FlhB function: The FlhB c-terminal  
170 homologue HP1575 acts as a "spare part" to permit flagellar export when the  
171 HP0770 FlhBCC domain is deleted. *J. Bacteriol.* **188**, 7531-7541 (2006).
- 172 19. H. E. Hoekstra, J. A. Coyne, The locus of evolution: evo devo and the genetics of  
173 adaptation. *Evolution* **61**, 995-1016 (2007).
- 174 20. G. A. Wray, The evolutionary significance of *cis*-regulatory mutations. *Nat. Rev.*  
175 *Genet.* **8**, 206-216 (2007).

- 176 21. S. B. Carroll, Evo-devo and an expanding evolutionary synthesis: a genetic  
177 theory of morphological evolution. *Cell* **134**, 25-36 (2008).
- 178 22. M. W. Silby, *et al.*, Genomic and genetic analyses of diversity and plant  
179 interactions of *Pseudomonas fluorescens*. *Genome Biol.* **10**, R51 (2009).
- 180 23. F. Van Immerseel, *et al.*, *Salmonella Gallinarum* field isolates from laying hens  
181 are related to the vaccine strain SG9R. *Vaccine* **31**, 4940-4945 (2013).
- 182 24. K. Rutherford, *et al.*, Artemis: sequence visualization and annotation.  
183 *Bioinformatics* **16**, 944-945 (2000).
- 184 25. D. B. Roche, M. T. Buenavista, S. J. Tetchner, L. J. McGuffin, The IntFOLD  
185 server: an integrated web resource for protein fold recognition, 3D model quality  
186 assessment, intrinsic disorder prediction, domain prediction and ligand binding  
187 site prediction. *Nucleic Acids Res.* **39**, W171-W176 (2011).
- 188 26. L. J. McGuffin, M. T. Buenavista, D. B. Roche, The ModFOLD4 server for the  
189 quality assessment of 3D protein models. *Nucleic Acids Res.* **41**, W368-W372  
190 (2013).
- 191 27. L. J. McGuffin, Intrinsic disorder prediction from the analysis of multiple protein  
192 fold recognition models. *Bioinformatics* **24**, 1798-1804 (2008).
- 193 28. D. B. Roche, M. T. Buenavista, L. J. McGuffin, The FunFOLD2 server for the  
194 prediction of protein–ligand interactions. *Nucleic Acids Res.* **41**, W303-W307  
195 (2013).
- 196 29. Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on  
197 the TM-score. *Nucleic Acids Res.* **33**, 2302-2309 (2005).

**Acknowledgments:**

TBT, LJJ, RWJ and MAB conceived and designed the study. TBT, GM and AA performed experiments. MWS and AHD performed independent lines of enquiry on Pf0-2x. DS conducted bioinformatics analysis of genome resequencing data, identified mutated genes and handled sequencing data. LM conducted the protein structure prediction and analysis. This work was supported by a Leverhulme grant to LJJ, MAB and RWJ, BBSRC grant BB/J015350/1 to RWJ, start-up funding from the University of York to MAB, Qassim University to AA, and Agriculture and Food Research Initiative Competitive Grant 2010-65110-20392 from the USDA's National Institute of Food and Agriculture, Microbial Functional Genomics Program to MWS. TBT, GM, LJJ, RWJ, MWS, DJS and MAB wrote the paper. We thank Graham Bell, Mark Pagel, Angus Buckling and James Moir for useful discussions; Peter Ashton for processing of microarray data; and Konrad Paszkiewicz and Exeter Sequencing Service facility and support from the following: Wellcome Trust Institutional Strategic Support Fund (WT097835MF), Wellcome Trust Multi User Equipment Award (WT101650MA) and BBSRC LOLA award (BB/K003240/1). Sequence data for genomic resequencing of AR2S and AR2F have been submitted to the SRA under accession numbers SRR1510202 and SRR1510203, respectively. The eArray design ID for the microarray is 045642. Microarray data have been submitted to the ArrayExpress database under accession number E-MTAB-2788 ([www.ebi.ac.uk/arrayexpress](http://www.ebi.ac.uk/arrayexpress)).

## Figures 1 – 3:

### Fig. 1. Phenotypic assays of motility variants.

**A** Surface spreading motility assays of ancestral (AR2) and evolved slow spreading (AR2S) and fast spreading (AR2F) mutants, after 27 hours. **B** Electron microscopy confirms the presence of a flagellum in AR2F, but fails to confirm presence in AR2S. **C** Mean (N=4) cell doublings per hour (+/- 1 SEM). Strains were grown in differing nitrogen environments: 10 mM glutamine (Gln), glutamate (Glu) and ammonium (NH<sub>4</sub>) as the sole nitrogen source, or in high nutrient lysogeny broth (LB): AR2,  $F_{3,12} = 13.460$ ,  $P < 0.001$ ; AR2S,  $F_{3,12} = 72.674$ ,  $P < 0.001$ ; AR2F,  $F_{3,12} = 52.538$ ,  $P < 0.001$ . There were also differences between doubling rate of strains within each nitrogen medium (Glutamate (Glu),  $F_{2,9} = 12.654$ ,  $P = 0.002$ ; Ammonium (NH<sub>4</sub>),  $F_{2,9} = 40.529$ ,  $P < 0.001$ ), with the exception of Glutamine (Gln) ( $F_{2,9} = 3.703$ ,  $P = 0.067$ ).

### Fig. 2. Heat map of microarray expression profiles for all evolved and ancestral motility variants.

Heat maps show where there is significant ( $P \leq 0.05$ ) fold-change of  $\geq 2$  in genes selected based on GO-terms for **A** Bacterial-type flagellum (24 genes) and **B** Nitrogen compound transport (146 genes) for all strains. The gradation of colors reflects normalized raw signal values across the entire array. Genes are ordered according to chromosomal position to enable clearer visualization of coregulated gene clusters. Full transcriptome analysis is reported in Supplementary File “Microarray dataset.xlsx”.

### Fig. 3. Full chain multi-template 3D models of protein structures of slow and fast spreading motility variants.

Slow spreading variants can either follow the direct regulatory route through mutation of NtrB or GlnK (**A**), or the physiological route through mutation of GlnA causing over-activation of NtrB (**B**); both routes are predicted to lead to hyperphosphorylation of NtrC. Fast spreading variants all show mutational changes to NtrC (**C**), which has a similar global structure to FleQ (**D**). The colour scheme represents the variation in models,

which correlates with local (per-residue) quality and disorder. Regions coloured in blue and green represent low local variability in structure, while those in red show high local variability (see Table 1 and Table S2 for mutation details). § = All mutations mapped onto SBW25 wildtype protein structures for illustrative purposes; \* = Truncated domain.

	<b>Slow Spreaders (AR2/Pf0-2xS)</b> Hyperphosphorylation of NtrC	<b>Fast Spreaders (AR2/Pf0-2xF)</b> Switch of NtrC-P specificity
<b>AR2</b>	<i>ntrB</i> T97P*	<i>ntrC</i> R442C*
	<i>ntrB</i> V185K	<i>ntrC</i> K342E
	<i>ntrB</i> D179N	<i>ntrC</i> G452R
	<i>ntrB</i> L184Q / V185I	<i>ntrC</i> K342V / Frameshift: V342
<b>Pf0-2x</b>	<i>ntrB</i> D228A <sup>§</sup>	<i>ntrC</i> N454S
		<i>ntrC</i> R441S
		<i>ntrC</i> N454S
		<i>ntrC</i> P424L
		<i>ntrC</i> N454S
	<i>ntrB</i> D228A <sup>§</sup>	<i>ntrC</i> L418R
		<i>ntrC</i> G414D
		<i>ntrC</i> N454S
		<i>ntrC</i> F426V
	<i>glnK</i> Frameshift: I86*	<i>ntrC</i> R442H*
<i>glnA</i> T237P*	<i>ntrC</i> A445V*	
<i>glnA</i> Frameshift: T205*	<i>ntrC</i> R442C*	

\* = Identified by genome resequencing

§ = Independent *ntrB* mutant strains, parent to multiple *ntrC* mutant strains

**Table 1. Mutational trajectory towards slow and fast spreading phenotypes.**

Mutations confirmed in slow spreading motility variants are predicted to result in hyperphosphorylation of NtrC; mutations in fast spreading variants lead to predicted switched specificity of NtrC-P towards FleQ targets. Slow and fast spreading variants share the same ancestry.



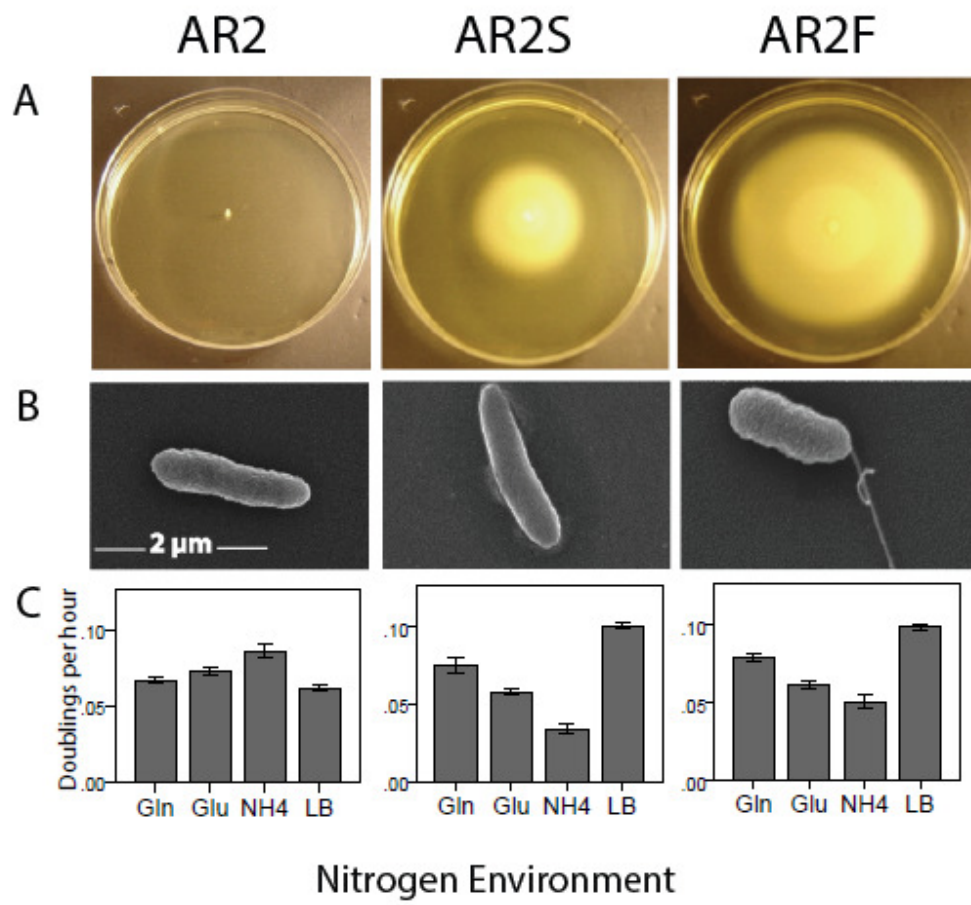


Figure 1

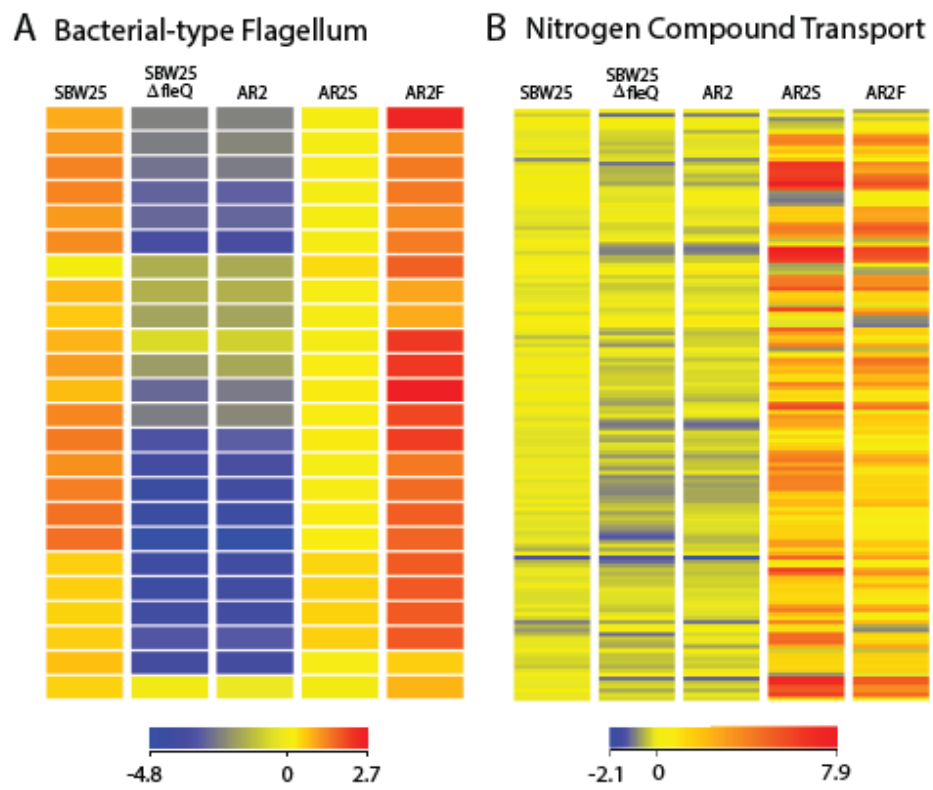


Figure 2

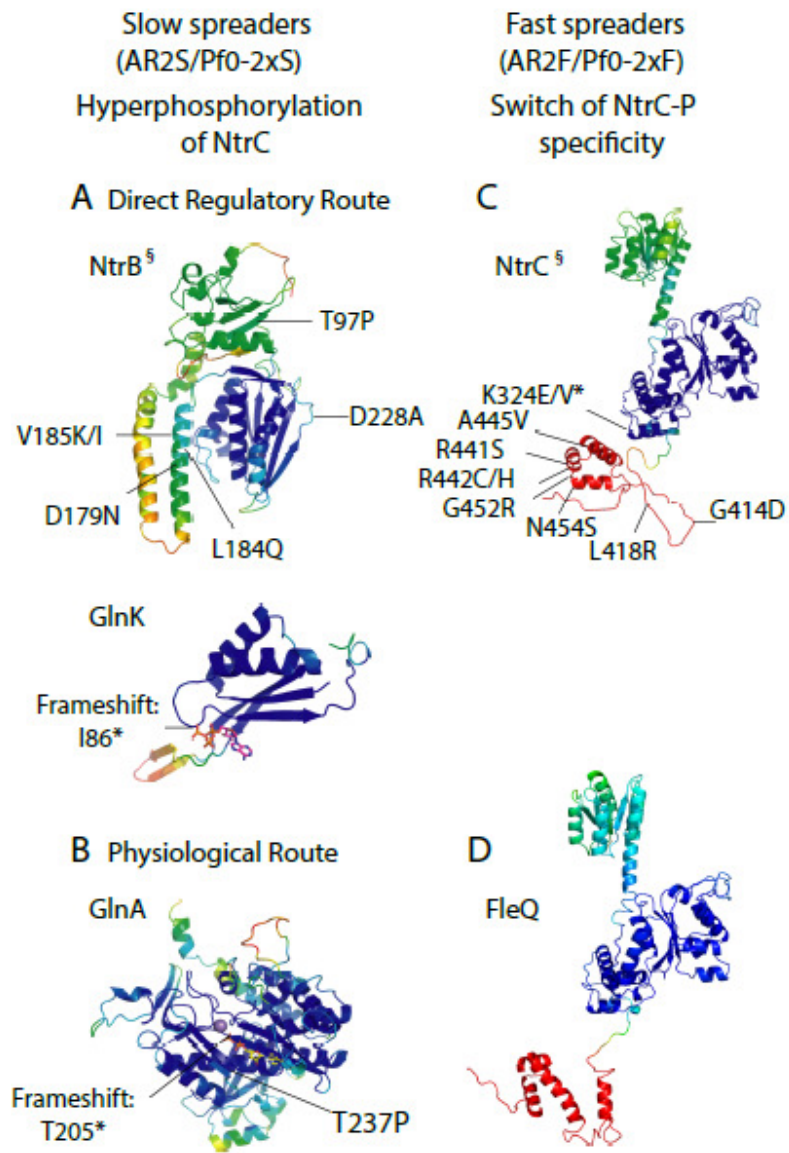


Figure 3

## Supplementary Materials:

Title: Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system

Authors: Tiffany B. Taylor<sup>1†</sup>, Geraldine Mulley<sup>1†</sup>, Alexander H. Dills<sup>2</sup>, Abdullah S. Alsohim<sup>1,3</sup>, Liam J. McGuffin<sup>1</sup>, David J. Studholme<sup>4</sup>, Mark W. Silby<sup>2</sup>, Michael A. Brockhurst<sup>5</sup>, Louise J. Johnson<sup>1,\*</sup>, Robert W. Jackson<sup>1,6</sup>

Materials and Methods

Fig. S1.

Table S1.

Table S2.

198 **Supplementary Materials:**

199

200 **Materials and Methods:**

201

202 **Microbiological methods**

203 *P. fluorescens* strains used in this study: SBW25, SBW25 $\Delta fleQ$ , AR2 (SBW25 $\Delta fleQ$  IS-  
204  $\Omega$ Km-hah: PFLU2552), and Pf0-2x (Pf0-1 $\Delta fleQ$ ). Motility assays are as described in  
205 (12). All starting populations were from a single AR2 colony grown on LB agar (1.5%),

206 and stab inoculated into the center of a SMM plate using a sterile wire. The initial  
207 observed motility mutants (AR2S and AR2F) were cultured immediately, cryopreserved  
208 and used for subsequent genome resequencing and microarray analysis. Independently  
209 evolved motility mutants were isolated from AR2 and Pf0-2x and cryopreserved.  
210 Mutation rates of all lineages were checked by plating an overnight culture onto LB agar  
211 supplemented with 100  $\mu\text{g ml}^{-1}$  rifampicin. All strains were found to have a similar  
212 mutation frequency: SBW25 =  $6.98 \times 10^{-9}$  cfu  $\text{ml}^{-1}$ ;  $\Delta\text{fleQSBW25}$  =  $3.72 \times 10^{-8}$  cfu  $\text{ml}^{-1}$ ;  
213 AR2 =  $2.50 \times 10^{-8}$  cfu  $\text{ml}^{-1}$ ; AR2S =  $2.3 \times 10^{-8}$  cfu  $\text{ml}^{-1}$ ; AR2F =  $1.4 \times 10^{-8}$  cfu  $\text{ml}^{-1}$ .  
214 M9 nitrogen modified medium (lacking ammonium) was supplemented with 10 mM  
215 glutamate, glutamine or ammonium solutions. Strains were grown for 16 hours in LB at  
216 27°C and diluted to OD 0.001 in M9 nitrogen modified media, and in LB. Optical density  
217 595 nm was measured every 20 minutes, under continuous shaking and at an  
218 incubation of 27°C, for 24 hours (Tecan, GENios).

219

## 220 **Molecular methods**

221 The mutations present in the original evolved mutants AR2S and AR2F, plus 3 Pf0-2xS  
222 and 3 descended Pf0-2xF strains were identified by genome resequencing.

223 Genomic DNA was isolated using Puregene DNA extraction (Qiagen) or Wizard®  
224 Genomic DNA Purification (Promega) kits. We resequenced the complete genomes of  
225 mutants AR2S and AR2F (original) using the Illumina HiSeq and identified single-  
226 nucleotide variants (SNV) by alignment to the SBW25 reference as described previously  
227 (22, 23). Further mutants were analyzed by targeted PCR amplification and sequencing

228 of *ntrBC* to determine whether the same two-step process had occurred; primers were  
229 designed to flank the region where SNVs had been identified by whole-genome  
230 resequencing (NtrC, F: CTTTCATCCCCAACTCCTTGA, R:  
231 AAGCTGCTGAAAAGCGAGAC; NtrB, F: CTTGCGCCTTGAGTACATGA, R:  
232 ATGCGGTCTACCAGGTTACG).

233

#### 234 *AR2S and AR2F Isolate Genome Resequencing Details*

235 Whole-genome resequencing was performed using the Illumina GA2x. We generated  
236 12,065,035 pairs of 36-bp reads for AR2S (SRA accession SRR1510202) and  
237 24,075,130 pairs of 36-bp reads for AR2F (SRR1510203). To identify mutations, we  
238 aligned the genomic sequence reads against the *Pseudomonas fluorescens* SBW25  
239 genome sequence (NCBI RefSeq accession NC\_012660) using BWA-mem version  
240 0.7.5a-r405 (<http://www.ncbi.nlm.nih.gov/pubmed/19451168>). We discarded reads that  
241 did not uniquely map to a single location on the SBW25 genome in order to avoid  
242 artifacts due to misalignment of sequence reads arising from repeat sequences. The  
243 resulting average depths of coverage over the SBW25 genome were 124x and 144x for  
244 AR2S and AR2F, respectively. For 99.86% of the SBW25 reference genome sequence  
245 (6,713,197 out of 6,722,539 bp) we were able to unambiguously determine the  
246 nucleotide sequence in both mutant genomes (on the basis that at these genomic  
247 positions there was at least 95% consensus among all aligned reads at those positions  
248 for each of the two resequenced genomes and depth of at least 5x). Over the 6,713,197  
249 bp of the genome for which we could unambiguously determine the DNA sequence for  
250 both mutant genomes, we found only three SNVs with respect to the reference genome

251 sequence. Two of these three variants were present in both resequenced genomes: at  
252 position 376,439 both resequenced strains had a G whereas in the SBW25 reference  
253 genome the base is T. This corresponds to a non-silent change from codon acc to  
254 codon Ccc in gene PFLU0344 (*glnL/ntrB*). The second variant was at position 1,786,536  
255 where SBW25 has A but the two resequenced mutants have G; this variant falls in an  
256 intergenic region and we have no reason to suspect it has an effect on the phenotypes  
257 observed. The third SNV is private to AR2F and falls within the gene PFLU0343  
258 (*glnG/ntrC*). In SBW25 and AR2S the base is G but in AR2F it is A; this results in a non-  
259 silent change from codon cgt to codon Tgt. Aside from these three variant sites, the  
260 remainder of the 6,713,197 bp of unambiguously resolved genome sequence contained  
261 no variation from the SBW25 reference genome.

262

### 263 *Pf0-2xS and Pf0-2xF Isolate Genome Resequencing Details*

264 Library preparation was performed using Illumina Nextera XT kit with Nextera XT  
265 Indexes and sequenced as 250PE reads from MiSeq.

266 Whole-genome resequencing of Pf0-2x strains was performed using the Illumina MiSeq.  
267 We generated 1,145,122 forward and 1,141,017 reverse reads of 251-bp for Pf0-2xS\_1,  
268 1,279,360 forward and 1,272,990 reverse reads of 251-bp for Pf0-2xF\_1.1, 1,962,339  
269 forward and 1,941,676 reverse reads of 251-bp for Pf0-2xS\_2, 907,328 forward and  
270 906,083 reverse reads of 251-bp for Pf0-2xF\_2.1, 935,862 forward and 932,415 reverse  
271 reads of 251-bp for Pf0-2xS\_3, and 1,194,665 forward and 1,190,975 reverse reads of  
272 251-bp for Pf0-2xF\_3.1. These genomic sequence reads were mapped against the  
273 *Pseudomonas fluorescens* Pf0-1 genome sequence (NCBI RefSeq accession

274 NC\_007492.2) using CLC Genomics Workbench 6.0; unmapped reads were not  
275 included in further analysis. The average coverage depths for each library were: 44.6 for  
276 Pf0-2xS\_1, 49.8 for Pf0-2xF\_1.1, 76.1 for Pf0-2xS\_2, 35.3 for Pf0-2xF\_2.1, 36.4 for Pf0-  
277 2xS\_3, and 46.5 for Pf0-2xF\_3.1. Out of the 6,438,405 bp Pf0-1 genome, 97.06-99.35%  
278 of nucleotides (depending on sample) had a minimum 95% consensus among all  
279 mapped reads. Using probabilistic variant detection to investigate unambiguous  
280 nucleotide positions in each sequenced strain, we identified two SNVs and a 3bp  
281 deletion common to all mutants (Table S2). These three common changes were also  
282 present in the parental strain (Pf0-2x) and were thus not considered further. A 15bp  
283 deletion was identified in both the Pf0-2xS1 and F1.1 strains with a unique SNV  
284 occurring in the latter. An additional SNV was common to Pf0-2xS2 and Pf0-2xF2.1, yet  
285 another to both Pf0-2xS3 and Pf0-2xF3.1, and a single unique SNV was discovered in  
286 Pf0-2xF2.1 and another in Pf0-2xF3.1 with respect to the reference genome sequence.

287  
288 The presence of *glnA*, *glnK* and *ntrC* mutations in resequenced strains was confirmed  
289 by PCR of candidate regions from parental and evolved strains followed by Sanger  
290 sequencing of the amplicons. Mutations in *ntrB* were detected by targeted PCR  
291 amplification and sequencing, and *ntrC* mutations in derivatives of *ntrB* mutants were  
292 detected the same way. In some cases, multiple primers were used to cover the full  
293 length of the gene. This enabled us to be certain about the mutations, and rule out any  
294 others. Primers used for Pf0-1 were: *ntrB*, F: ATTGCGCCTCGAGTACATGA and  
295 TCGACACGGTTTCACTACGG, R: ATGCGGTGACAGATTGCG and  
296 TCGGAGCCTTGGTTTGGTTT; *ntrC*, F: AAGCTGCTGAAAAGCGAGAC and



297 GATTAAGGGTCACGGTGCCT, R: CTTCATGCCGAGTTCCTTGA and  
298 CACTGGAACAAGGAGCCACA; glnA, F: GGAGGCCTTTCTTTGTCACG and  
299 ACGCTTGTAGGAGTTGGTGG and CGGGAAGTAGCCACCTTTCA, R:  
300 CCACCAACTCCTACAAGCGT and CAAGTCCGACATCTCCGGTT and  
301 CTACCCGCCCTAATTCACCC; glnK, F: GCCGGGCATTACGATAGACA, R:  
302 TGCCTTAGACTTGAGTCGG.

303

#### 304 *Microarray design*

305 For microarray analysis, RNA was extracted from SBW25, SBW25 $\Delta$ *fleQ*, AR2, AR2S  
306 and AR2F using an RNeasy kit (Qiagen) and assayed for quality [2100 Bioanalyzer  
307 (Agilent), and Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific)]. RNA was  
308 labelled, hybridised to custom Agilent arrays and scanned according to the  
309 manufacturer's instructions (Technology Facility, Dept. of Biology, University of York,  
310 UK). Differentially expressed genes were identified by ANOVA using the Benjamini-  
311 Hochberg FDR correction. No commercially available microarrays were available for *P.*  
312 *fluorescens*, so a custom design was created using the Agilent eArray system  
313 (<http://earray.chem.agilent.com/earray>). The *P. fluorescens* SBW25 genome was  
314 loaded into Artemis (24) (<http://www.sanger.ac.uk/resources/software/artemis>), and all  
315 open reading frames marked as CDSs and these putative CDS written to a FASTA file  
316 which was uploaded to eArray. A probeset was created containing five unique probes  
317 per CDS, plus the appropriate Agilent control sequences. The probes were then laid out  
318 onto a standard Agilent 8x60K format slide. Microarrays were validated using qRT-PCR  
319 (MyiQTM, Bio-Rad).

320

321 *RNA labelling and hybridisation*

322 RNA isolated from each sample was labelled with Cy-3 using the Agilent Low Input  
323 Quick Amp WT Labelling Kit (one color) according to the manufacturer's instructions.  
324 Briefly, the RNA is reverse transcribed using a primer mix containing oligo-dT and  
325 random nucleotide primers containing T7 promoters, then the cRNA amplified using the  
326 T7 polymerase in the presence of Cy-3 labelled dCTP. The labelled cDNA was then  
327 quality controlled and only samples with a specific activity of > 6 were used hybridised  
328 to the arrays.

329 Samples were hybridised to the arrays, incubated and washed in accordance with the  
330 manufacturer's instructions, and the slides scanned using an Agilent Microarray  
331 Scanner and the fluorescence quantified with the Feature Extraction software. The  
332 resulting files were then loaded into GeneSpring for further analysis.

333

334 *Initial Data Analysis*

335 GeneSpring software version 12.6 was used for all data analysis. Initial analysis was  
336 performed to identify those genes that were differentially expressed between any  
337 conditions in the experiment, followed by Gene Set Expression Analysis (GSEA) on the  
338 basis of Gene Ontology terms. Briefly, the sample replicates were grouped, and then  
339 ANOVA performed to identify differentially expressed genes, using the Benjamini-  
340 Hochberg FDR p-value correction (with a FDR of 5%), and a Tukey HSD post-hoc test  
341 used for each gene to identify which samples were significantly different from each

342 other for that gene. Those GO categories that were over-represented in the set of  
343 differentially expressed genes were then identified.

344

### 345 **3D protein models**

346 The IntFOLD server (25) version 2 was used to build multi-template 3D models from the  
347 wild type and mutant sequences of AdnA, NtrB, NtrC, FleQ, GlnA and GlnK. The 3D  
348 models were quality checked using the ModFOLD4 protocol (26). Intrinsic protein  
349 disorder was predicted using the DISOclust method (27). Binding sites were predicted  
350 using the FunFOLD2 method (28). Models were structurally aligned and scored using  
351 TM-align (29).

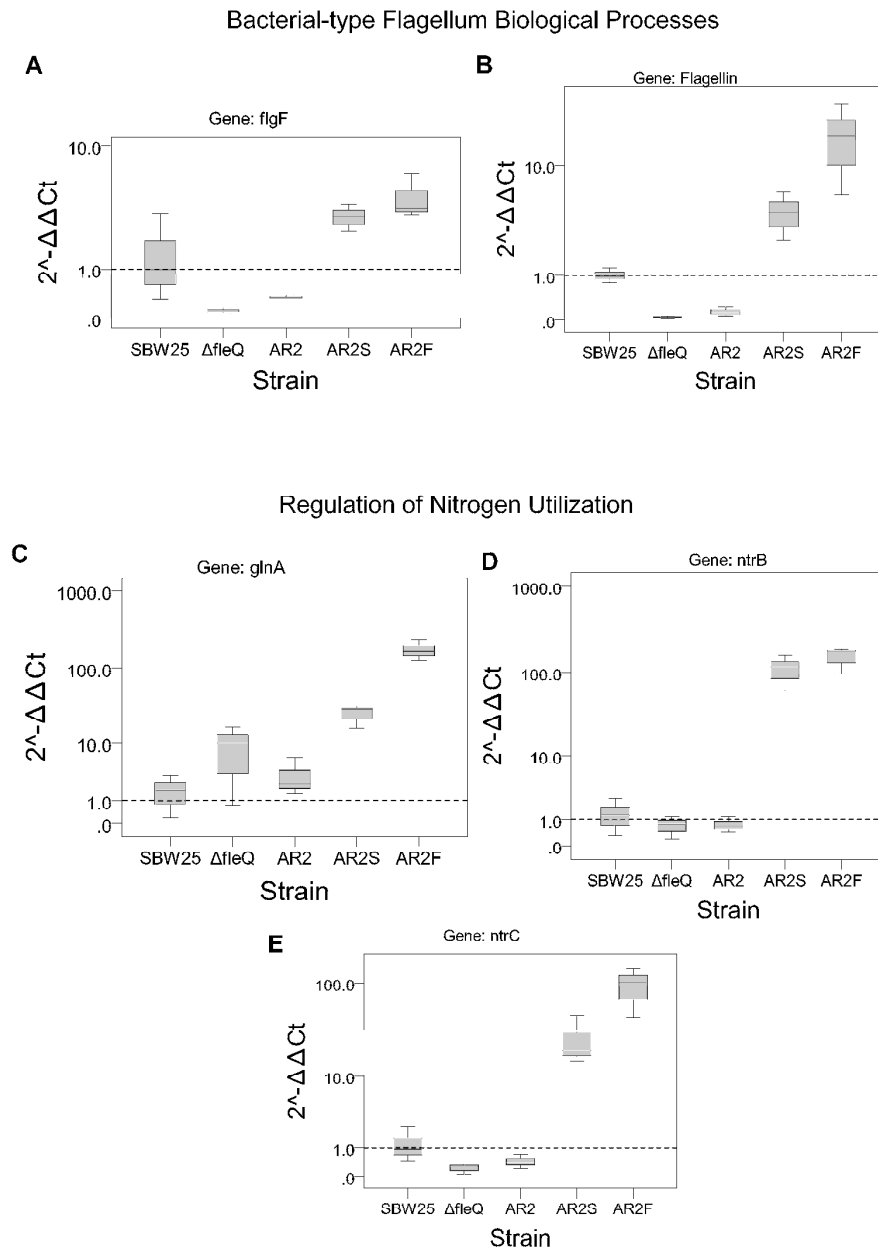
352

### 353 **Further Discussion of Microarray Results**

354 Full transcriptome analysis is reported separately in Supplementary File “Microarray  
355 dataset.xlsx”. When genes were filtered for significant fold changes ( $P \leq 0.05$ ) of  $\geq 2$ ,  
356 and sorted by descending fold changes, the genes that showed the greatest changes in  
357 expression are those highlighted in this study. Specifically, when comparing AR2 and  
358 AR2S, the large majority of genes with greatest up regulation are related to nitrogen  
359 transport, whereas when comparing AR2S to AR2F, the large majority of genes are  
360 related to flagellar motility and chemotaxis.

361 Our microarray data show that the average level of flagella gene expression in AR2S is  
362 between 1.5-2-fold lower than in SBW25, but considerably higher than AR2 (between 3  
363 and 75 fold). There are at least two possible models to explain why we were unable to

364 observe flagella in AR2S: (i) Phosphorylated NtrC (NtrC-P) only weakly interacts with  
365 flagella gene promoters, but the over-abundance of NtrC-P in AR2S saturates NtrC-  
366 dependent promoters and the excess NtrC-P drives expression of flagella genes. The  
367 level of transcription is suboptimal, thus the flagella may be unstable and support limited  
368 or transient motility; (ii) Promiscuous activity of over-abundant NtrC-P (e.g. interacting  
369 with RpoN from solution) leads to expression from flagella promoters, but this is not  
370 equivalent in all cells and only a proportion of the population express sufficient levels to  
371 produce a functional flagellum at any one time.



372

**Fig. S1: Validation of gene expression changes detected by microarray using qRT-PCR.**

**A – E** Box plots show gene expression, relative to wild-type SBW25, in genes related to nitrogen utilization and flagellar function. Note, data presented on log scales.

A. SBW25Δ*fleQ*

	GENE SYMBOL	GENE CHIP ARRAY		REAL TIME PCR	
		Fold Change	p-value	Fold Change	p-value
Bacterial-type Flagellum	<i>flgF</i>	-79.06	1.52E-07	-8.4	0.196
	Pflu4448 (Flagellin)	-158.33	2.09E-07	-22.59	0.002
Regulation of Nitrogen Utilization	<i>glnA</i>	1.01	2.24E-06	9.14	0.187
	<i>ntrB</i>	-1.54	1.37E-07	-4.5	0.41
	<i>ntrC</i>	-1.74	1.58E-07	-1.39	0.136

## B. AR2

	GENE SYMBOL	GENE CHIP ARRAY		REAL TIME PCR	
		Fold Change	p-value	Fold Change	p-value
Bacterial-type Flagellum	<i>flgF</i>	-76.48	1.52E-07	-2.84	0.265
	Pflu4448 (Flagellin)	-128.22	2.09E-07	-6.61	0.004
Regulation of Nitrogen Utilization	<i>glnA</i>	1.03	2.24E-06	3.27	0.393
	<i>ntrB</i>	-1	1.37E-07	-2.24	0.424
	<i>ntrC</i>	-1.41	1.58E-07	-1.31	0.231

## C. AR2S

	GENE SYMBOL	GENE CHIP ARRAY		REAL TIME PCR	
		Fold Change	p-value	Fold Change	p-value
Bacterial-type Flagellum	<i>flgF</i>	-2.53	1.52E-07	3.13	0.192
	Pflu4448 (Flagellin)	-1.71	2.09E-07	4.34	0.042
Regulation of Nitrogen Utilization	<i>glnA</i>	6.08	2.24E-06	25.49	0.008
	<i>ntrB</i>	28.18	1.37E-07	26.45	0.017
	<i>ntrC</i>	27.14	1.58E-07	49.14	0.059

## D. AR2F

	GENE SYMBOL	GENE CHIP ARRAY		REAL TIME PCR	
		Fold Change	p-value	Fold Change	p-value
Bacterial-type Flagellum	<i>flgF</i>	1.1	1.52E-07	4.49	0.1
	Pflu4448 (Flagellin)	3.6	2.09E-07	16.57	0.065
Regulation of Nitrogen Utilization	<i>glnA</i>	11.34	2.24E-06	175.63	0.005
	<i>ntrB</i>	58.27	1.37E-07	96.09	0.006
	<i>ntrC</i>	63.72	1.58E-07	49.73	0.03

**Table S1: Mean fold change (relative to SBW25) from qRT-PCR and microarrays with p-values (A – D).**

Colours correspond to genes with significant ( $\geq 2$  fold change) up regulated (red) and down regulated (blue) changes.

Strain	Nucleotide change	AA change	Gene	Domain/Function
AR2S_1	T376439G*	T97P	<i>ntrB</i>	PAS domain
AR2F_1.1	G374322A*	R442C	<i>ntrC</i>	HTH domain
AR2S_2	G376175A/ T376174A	V185K	<i>ntrB</i>	HK domain
AR2F_2.1	T374622C	K342E	<i>ntrC</i>	HTH domain
AR2S_3	G376193A	D179N	<i>ntrB</i>	HK domain
AR2F_3.1	C374291G	G452R	<i>ntrC</i>	Miscellaneous function
AR2S_4	G376176A T376174A	L184Q V185I	<i>ntrB</i>	HK domain
AR2F_4.1	T374622C/ /A374625	K342V FS from V342	<i>ntrC</i>	HTH domain
Pf0-2xS_1	$\Delta$ 6431771-3*	$\Delta$ G85	<i>gidB</i>	Methyltransferase domain
	$\Delta$ 6171472-87*	FS from I86	<i>glnK</i>	PII regulator
Pf0-2xF_1.1	C382879T*	R442H	<i>ntrC</i>	HTH domain
Pf0-2xS_2	$\Delta$ 6431771-3*	$\Delta$ G85	<i>gidB</i>	Methyltransferase domain
	T388861G* $\Delta$ 902225*	T237P FS from S544	<i>glnA</i> <i>carB</i>	Catalytic domain Miscellaneous function
Pf0-2xF_2.1	G382870A*	A445V	<i>ntrC</i>	HTH domain
Pf0-2xS_3	$\Delta$ 6431771-3*	$\Delta$ G85	<i>gidB</i>	Methyltransferase domain
	$\Delta$ 388958*	FS from T205	<i>glnA</i>	Catalytic domain
Pf0-2xF_3.1	G382880A*	R442C	<i>ntrC</i>	HTH domain
Pf0-2xS_4	A384603C	D228A	<i>ntrB</i>	Miscellaneous function
Pf0-2xF_4.1	A382843G	N454S	<i>ntrC</i>	HTH domain
Pf0-2xF_4.2	C382883A	R441S	<i>ntrC</i>	HTH domain
Pf0-2xF_4.3	A382843G	N454S	<i>ntrC</i>	HTH domain
Pf0-2xF_4.4	C382933T	P424L	<i>ntrC</i>	HTH domain
Pf0-2xF_4.5	A382843G	N454S	<i>ntrC</i>	HTH domain
Pf0-2xS_5	A384603C	D228A	<i>ntrB</i>	Miscellaneous function
Pf0-2xF_5.1	T382951G	L418R	<i>ntrC</i>	Miscellaneous function
Pf0-2xF_5.2	G382963A	G414D	<i>ntrC</i>	Miscellaneous function
Pf0-2xF_5.3	A382843G	N454S	<i>ntrC</i>	HTH domain
Pf0-2xF_5.4	T382928G	F426V	<i>ntrC</i>	HTH domain

\* = Identified by genome resequencing.

**Table S2: Full details of mutations identified in strains with parental AR2 and Pf0-2x origin.**

Numbers before decimal points represent independent lineages, and numbers after represent derived strains. Note fast spreading mutants contain all nucleotide changes within a lineage.