Evolutionary innovation through transcription factor promiscuity in microbes is constrained by pre-existing gene regulatory network architecture

Authors: Matthew J. Shepherd[1], Aidan P. Pierce[1], Tiffany B. Taylor[1]*

1 - Milner Centre for Evolution, Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK

* - Corresponding author: Tiffany B. Taylor, Milner Centre for Evolution, Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK, +44 (0)1225 384398, t.b.taylor@bath.ac.uk

The contributions to this work were as follows:

Experimental work performed by MJS. Rhamnose expression plasmids built by APP. Manuscript written by MJS. Manuscript editing by all authors. Project supervised by TBT. Funding secured by TBT.

Acknowledgements:

This work was funded by a Royal Society Research Fellows Grant (RG160491; awarded to TBT) supporting MJS; Royal Society Research Fellows Enhancement Award (RGF\EA\201057; awarded to TBT) supporting APP; a Royal Society Dorothy Hodgkin Research Fellowship (DH150169; awarded to and supporting TBT).

Bioinformatics analyses for the article was carried out using the Medical Research Council's (MRC) Cloud Infrastructure for Microbial Bioinformatics (CLIMB), Illumina Whole-Genome Sequencing conducted by MicrobesNG, Birmingham, UK (http://www.microbesng.com) which was supported by the BBSRC (grant number BB/L024209/1), and Illumina RNA sequencing with ribodepletion was conducted by Oxford Genomics Centre, Oxford UK.

We would like to thank Prof. Laurence Hurst for insightful discussion of earlier manuscript versions, Dr. Jonathan Nzakizwanayo and Dr. Iain MacArthur for providing us with *E.coli* strains SP50 pRK2073, S17 pTNS2 and ST18, Dr. Ellie Harrison for providing the R script that Fig.4A is based upon, and Dr. James Horton, Louise Flanagan and Josie Elliott for helpful discussion during the course of the study.

Abstract

The survival of a population during environmental shifts depends on whether the rate of phenotypic adaptation keeps up with the rate of changing conditions. A common way to achieve this is via change to gene regulatory network (GRN) connections conferring novel interactions on transcription factors. To understand the success of rapidly adapting organisms, we therefore need to determine the rules that create and constrain opportunities for GRN innovation. Here, using an experimental microbial model system we construct a maladapted GRN, through deletion of a master transcription factor, and challenge evolution to fix this network. We identify three key properties - high activation, high expression, and pre-existing low-level affinity for novel target genes – that facilitate transcription factor innovation via gain of functional promiscuity. Ease of acquiring these properties is constrained by pre-existing GRN architecture, which was overcome in our experimental system by both targeted and global network alterations. This work reveals the key properties that determine transcription factor evolvability, and as such, the evolution of GRNs.

Introduction

During conditions of environmental upheaval and niche transition events, rapid phenotypic adaptation is essential for evolutionary success[1–3]. To understand patterns of diversification in novel environments we need to understand why some evolutionary transitions occur more rapidly than others, and what allows some organisms to succeed where others fail. Key to this is understanding the evolution of gene regulatory networks (GRNs), control circuits common throughout the domains of life that determine the magnitude and timing of gene expression[4] in response to environmental and internal signals[5,6]. GRNs are frequent sites of adaptive mutation driving phenotypic evolution[7–9], and often underscore adaptation to - and survival in - new and changeable environments[10–13]. Determining how these networks evolve and identifying rules governing when and how a regulatory circuit adapts will allow us to better understand their role in determining the evolutionary success of an organism.

Within GRNs, a key mechanism of innovation is transcription factor rewiring, in which transcription factors can gain or lose regulatory connections to target genes creating new network architectures and opportunity for phenotypic innovation[14]. Rewiring events can have dramatic effects on the transcriptome[15] that can drive phenotypic diversification. However, the majority of past studies on transcription factor rewiring involve retrospective experimental dissection of networks that have already diverged[16–21], which allows inference of past rewiring events but does not address the evolutionary factors driving the rewiring process. For rewiring to occur, a transcription factor must first gain promiscuous activity allowing regulation of a new gene[22], as is the case for innovation in other proteins[23,24]. Functional

promiscuity constitutes low-affinity regulatory interactions that can be said to be non-cognate for the transcription factor in question; interactions that do not elicit a gene expression response[25] or elicit a response that has no physiological significance to the organism under normal conditions[26]. It is precisely these non-cognate interactions that can become meaningful and drive adaptation if favoured by natural selection[27,28]. However, promiscuity cannot be commonplace within GRNs as this will likely result in dysregulation of gene expression and fitness costs for an organism in the environment to which it is adapted. Indeed mechanisms exist to dampen promiscuity and maintain precise regulation within networks[29,30].

Studies on the evolution of promiscuity in transcription factors are rare. Primarily, investigations have been retrospective analyses focussing on examples of promiscuity driving transcription factor innovation in evolutionary history, such as the case of *MALR* regulators in *S. cerevisiae*[31], or the LEAFY[32] and MADS-domain transcription factors[33] in plants. A pair of bacterial experimental studies observing rewiring of transcription factors also observed probable promiscuity. These studies challenged evolutionary rescue of a motility trait lost due to deletion of a master transcription factor, and observed rewiring of the transcription factors NtrC and NmpR to restore motility[34,35]. In both cases, initial rescue of motility occurred without mutation to the DNA binding domain of the rewired transcription factors, indicating a gain of promiscuity. These transcription factors are both homologs of the respective missing master regulators, indicating that to regulate a non-cognate gene a degree of structural similarity to that genes' typical regulator is important. These transcription factors are both homologs of the respective missing master regulators, indicating that for a non-cognate interaction a degree of structural similarity to a genes cognate regulator is important.

These highlighted studies provide examples of promiscuity and its importance for driving evolutionary innovation in transcription factors. However, in each case it is unknown why the transcription factor in question became promiscuous as opposed to any of the other regulators within their GRN; what role – if any – GRN structure played in the process; and whether the rewired transcription factors were unique in their ability to become promiscuous within each study system. To address these questions, we set out to investigate the evolution of promiscuity using the rewiring of NtrC documented by Taylor et al. (2015b) as a model system. This system is based on *Pseudomonas fluorescens* AR2, a Δ*fleQ*, -*viscB* variant of strain SBW25 that lacks the flagellar master regulator FleQ and the ability to produce the biosurfactant viscosin. Rendering AR2 immotile on soft agar[34], these manipulations create a strong selective pressure for rescue of flagellar gene expression to restore motility in this environment. This provides a model of a maladapted GRN, which can

be used to study regulatory innovation in a repeatable experiment challenging evolution to 'fix' the network. AR2 was found to evolve to rescue flagellar motility through co-option of a FleQ-homologous transcription factor, NtrC. This occurred as a two-step process; an initial motility-granting mutation to the gene *ntrB* encoding NtrC's cognate kinase and a secondary motility-enhancing mutation to the helix-turn-helix (HTH) DNA binding domain of the NtrC transcription factor. These mutations allowed neofunctionalization and subsequent refinement of NtrC acting to induce flagellar gene expression, despite the NtrBC system typically regulating nitrogen assimilation[36]. This evolutionary rewiring to rescue flagellar gene expression exclusively occurred through the NtrBC system despite there being 20 other FleQ structural homologs present in the SBW25 genome (a transcription factor family called RpoN-dependent enhancer binding proteins – RpoN-EBPs), many of which are predicted to be more structurally similar to FleQ than NtrC[37]. This evolutionary 'preference' for innovation through the NtrBC system over any other FleQ-homologous regulator persists even after disruption of a strong mutational hotspot present in the *ntrB* gene[38,39]. Across these studies, a total of 387 independent replicates of motility rescue were run including varying nutrient conditions, and in two different strain backgrounds (AR2 and Pf0-2x, an engineered aflagellate Pf0-1 mutant), and in all cases mutations occurred in *ntrB*, or genes *glnK* and *glnA* that form part of the NtrBC system.

It is known that within families of homologous transcription factors there is variation in the ability to bind non-cognate sites[40], so it is possible that NtrBC is unique in its ability to become promiscuous and rewire. To test this, we construct an *ntrC* knockout in AR2 and placed this mutant under strong selection for swimming motility. This forces evolutionary utilisation of an alternative transcription factor in order to rescue motility, and is an established method for unveiling hidden evolutionary pathways[41]. Identification of additional transcription factors capable of gaining promiscuity within the same GRN background allows us to establish rules governing when and where a transcription factor becomes promiscuous within its GRN, and is important for understanding how these regulatory systems innovate during adaptation to environmental challenges.

Results

Evolutionary rescue of flagellar motility can occur in the absence of *ntrC*, through *de-novo* gain-of-function mutation to an alternative two-component system

To identify other RpoN-EBPs capable of promiscuity, *P. fluorescens* AR2Δ*ntrC* was challenged to rescue flagellar motility in 0.25% agar LB plates. Motile zones re-emerged in a two-step manner (a slow-swimming variant, then a faster-swimming variant). Motile isolates were sampled and whole genome re-sequenced at each step. Motility-granting mutations were identified in the gene *PFLU1131* for all 1st step motile isolates of AR2Δ*ntrC* (n=15), and in 13/15 cases this was the only mutant gene (Supplementary excel file E1). *PFLU1131* is unstudied and encodes a putative sensor histidine kinase. The gene is situated in an operon between two other genes, *PFLU1130* and *PFLU1132* which encode a hypothetical GNAT-acetyltransferase, and a putative sigma-54-dependent transcriptional regulator respectively. PFLU1132 is a known FleQ-homolog[37] and together with PFLU1131 forms a putative two-component system. The most frequent mutation in *PFLU1131* was an identical 15bp deletion (Fig.1A) in the histidine-kinase phospho-acceptor domain (73%). This mutation results in loss of amino acids 368-GEVAM-372 in the protein product (Henceforth referred to as PFLU1131-del15). Other mutations included a highly similar 15bp deletion resulting in loss of amino acids 369-EVAMG-373, as well as a SNP resulting in A375V. All of these mutations (86% of the 1st step mutations) cluster to the same 26bp of the 1770bp *PFLU1131* ORF, and result in amino acid changes in a site directly adjacent to the catalytically active H-box[42] (amino acids 376-382, Fig.1A) suggesting a gain-of-function effect on the putative kinase.

To confirm that observed motility phenotypes were dependent on the PFLU1131/2 two-component system, the *PFLU1132* gene was deleted and complemented with and without the presence of a 1st step kinase mutation (*PFLU1131*-del15). Knockout of *PFLU1132* abolished flagellar motility, and complementation restored motility only in the presence of the kinase mutation (Fig.1B). This experiment was repeated for *ntrBC* and produced the same result (Supplementary Fig.S1). The *PFLU1131*-del15 mutation also grants flagellar motility when engineered into AR2, so does not depend on the absence of *ntrC* to function (Supplementary Fig.S2). Transcriptomic analysis of the *PFLU1131*-del15 mutant by RNA sequencing indicates a gain-of-function effect on the PFLU1131/2 two-component system and a net up-regulatory effect on the transcriptome (Fig.1C). These regulatory changes include upregulation of the flagellar genes (Supplementary Fig.S3) and the *PFLU1130/1/2* operon.

The 1st step PFLU1131 mutations grant promiscuous activity to the PFLU1132 transcription factor

In 13/15 1st step mutants, only a single mutation to PFLU1131 was identified. To confirm that no other mutations were needed for flagellar motility, the *PFLU1131*-del15 mutation was introduced into the AR2ΔntrC ancestor. The resulting engineered strain was motile (Supplementary Fig.S2). As no mutational change to the DNA binding domain of PFLU1132 is needed, this suggests a non-specific mechanism through which PFLU1132 induced the flagellar genes.

To investigate this, our transcriptomic data was used to assess the impact of the *PFLU1131*-del15 mutation on the expression of a list of genes[43] controlled by all 22 RpoN-EBP's present in *P. fluorescens* SBW25[44]. This mutation resulted in significant upregulation for 54% of all RpoN-EBP controlled genes (Fig.2). This was less than the 70% of RpoN-EBP controlled genes upregulated in the NtrB T97P mutant reported previously[34]. Flagellar genes account for 12% of all RpoN-EBP controlled genes, and whilst the native regulatory targets of PFLU1132 are unknown it is unlikely that it natively regulates 54% of these. Similarly, the NtrB-T97P mutant reported in Taylor et al., (2015b) leads to upregulation of genes known to be involved in nitrogen assimilation, however it also results in upregulation of many genes with no known regulatory link to NtrC. The large percentages of the RpoN regulon induced by these specific kinase mutations indicate a gain of promiscuous regulatory activity in the RpoN-EBPs PFLU1132 and NtrC, which may be due to hyperphosphorylation and hyperactivation via mutant cognate kinases PFLU1131 and NtrB.

Evolutionary rescue of motility through PFLU1132 is significantly constrained in our assay

Time to emergence (the length of time taken for motility to evolve on the soft agar plate) was recorded for all replicate experimental lines of the AR2Δ*ntrC* ancestor, as well as AR2 as a comparison. Initial 1st step motile AR2Δ*ntrC* strains evolved with an average of 18.7 days from the assay start (Fig.3A), significantly longer than AR2 took to rescue motility (4.2 days, P = 0.002, Dunn test). After 1st step PFLU1131 mutations, 2nd step mutants evolved rapidly within an average of 2.2 days (Fig.3A). This was faster than the 2nd step mutants in the AR2 background which took 3.3 days to emerge (*P* = 0.0017, Dunn test). Importantly, only 9.4% of replicate experimental lines (n=160) of AR2Δ*ntrC* evolved within the 6 week assay cut-off compared to 100% for the AR2 background (n=22).

The greatly increased time taken to rescue motility in an AR2ΔntrC background may be reflective of a small pool of possible PFLU1131 mutations that grant promiscuity to

PFLU1132. One replicate line gained a *de novo* mutation in *mutS* (G79.fs, Supplementary excel file E1), resulting in a frame shift and probable loss of function for the DNA mismatch-repair protein that it encodes[45]. This strain, along with strains derived from it possessed large numbers of additional SNPs (>80) including SNP A375V in PFLU1131. Loss of function to *mutS* is a known strategy to enhance mutability in *Pseudomonads*[46], and its occurrence may increase access to a motility-rescuing mutation in PFLU1131.

From a phenotypic perspective, the 1st step mutations in the *PFLU1131* pathway provide a much poorer motility phenotype than the analogous 1st step *ntrB* mutations (Fig.3B). In a race assay, AR2ΔntrC *PFLU1131*-del15 mutants only swam 0.31 mm for every 1 mm swam by AR2 ntrB-T97P (P=0.000356, Wilcox test). One possible explanation for this poor motility is a lack of significant upregulation for the flagellar filament subunit FliC (Supplementary Fig.S3) in the *PFLU1131*-del15 mutant. This PFLU1131 mutation does not provide any measurable pleiotropic cost in LB media (Fig.3C). AR2ΔntrC *PFLU1131*-del15 produces an average relative area under the growth curve (rAUGC) in LB broth of 1.55 which did not significantly differ for the average rAUGC of 1.48 for the AR2ΔntrC ancestor (P = 0.1132, Dunn test). The poor motility phenotype provided by *PFLU1131*-del15 suggests that PFLU1132 is only capable of weakly activating flagellar motility in this mutant and that its ability to act promiscuously is constrained by another factor.

We additionally constructed a PFLU1132 knockout to identify any further RpoN-EBPs capable of promiscuity. AR2ΔntrCΔPFLU1132 failed to rescue flagellar motility within the 6-week assay cut-off (Supplementary excel file E2, n= 192), suggesting evolutionary gain of promiscuity in other transcription factors may not be easily achieved under the conditions we tested.

Motility enhancing 2nd step mutations suggest alterations to global gene regulatory network can facilitate promiscuity in PFLU1132

Following the 1st step mutations in PFLU1131 that unlock flagellar motility, our motile isolates develop 2nd step mutations that act to boost motility speed (Supplementary Fig.S4A). These second step mutations offer clues to the nature of constraining factors limiting innovation through PFLU1131/2, as they represent evolutionary solutions to the poor motility provided to the 1st step mutations. Whole genome resequencing identified a diverse set of motility enhancing 2nd step mutations (Fig.4A). In contrast to the frequent 2nd step mutations observed in the HTH DNA-binding domain of NtrC[34], no analogous mutations were observed in the same domain of PFLU1132. Mutations that did occur in the PFLU1132 gene were parallel and identified in only 2 lines, impacting the receiver domain of the transcription factor (D111G and D111N). One line gained a secondary PFLU1131 mutation, a SNP resulting in

S555N in the HATPase domain along with the 1st step SNP A547D already present (Fig.1B), which may further boost kinase activity. A mutation to the promoter region of the PFLU1131/2 operon was also observed.

Other secondary mutations can be grouped into 3 other broad categories. The first are mutations in the operon PFLU1583 and PFLU1584, accounting for 24% of secondary mutations (Fig.4A, Purple). These encode a putative anti-sigma factor PP2C-like phosphatase and a putative STAS family anti-sigma factor antagonist respectively. Mutations in PFLU1583 were generally loss of function (e:g: Q77*, E144.fs, R198.fs) suggesting loss of repression on an unknown sigma factor. *P. fluorescens* SBW25 encodes ~31 putative sigma factors[44], PFLU1583/4 may act on RpoN, the partner sigma factor of PFLU1132 in regulating gene expression, however this cannot be tested easily through *rpoN* knockout. Transcriptomic analysis of the PFLU1583 Δ48-74 mutation identified sigma factor RpoE as upregulated, however *rpoE* knockout did not negate the motility enhancing effect of the mutation (Supplementary Fig.S5A). To test that PFLU1583/4 are not acting through another RpoN-EBP or a separate mechanism, we engineered PFLU1583 Δ48-74 in AR2ΔntrC in the absence of a PFLU1131 mutation. This strain was immotile, indicating that the PFLU1583/4 mutations act on PFLU1131/2 (Supplementary Fig.S5B).

The second category of mutations impact core gene expression components that will affect global GRN function (Fig.4A, Yellow). These included mutations to *rho* (T338I), *rpsK* (R93C) and *rpoC* (G1362.fs), that encode core gene expression machinery and will impact expression of most genes in the genome. The final category of motility enhancing mutations fit a metabolic theme (Fig.4A, blue). These mutations occur in genes involved in central carbon metabolism including the Embden-Meyerhoff-Parnas pathway, Pentose-phosphate pathway and the Krebs cycle, which will likely have a significant impact on global gene expression through alterations to central carbon flux.

The diversity and nature of these 2nd step mutations suggest global regulatory strategies to facilitate flagellar gene expression through PFLU1131/2. To understand their regulatory impact, transcriptome analysis was performed on a pair of representative mutants. The PFLU1583 Δ48-74 mutant was chosen due to the frequent mutation of PFLU1583 in our experiments, and the PFLU1131/2 promoter mutation as a representative of mutation specific to the PFLU1131/2 locus. Principle component analysis indicates that the transcriptomic profile of PFLU1583 Δ48-74 differs significantly from the profile of the PFLU1131/2 promoter mutation. Both mutations resulted in similar variation across PC1 relative to the 1st step *PFLU1131*-del15 mutant, but opposite variation in PC2 (Fig.4B). Both the PFLU1131/2 promoter and PFLU1583 Δ48-74 mutations result in net up-regulatory

effects on the transcriptome reflected by positive skewed volcano plots (Fig.4C and Fig.4D) albeit with differing expression patterns. Both mutations have the effect of further upregulating RpoN-EBP controlled genes with 22% and 60% being upregulated for PFLU1583 Δ48-74 and the PFLU1131/2 promoter mutant respectively (Fig.4E). This indicates that both mutations act to increase promiscuous regulation of the flagellar genes by PFLU1132, albeit through differing mechanisms.

2nd step PFLU1131/2 promoter capture mutation suggests importance of transcription factor expression in gaining promiscuity

One possible limitation on evolutionary innovation through PFLU1131/2 is the expression level of the two-component system itself. For promiscuous activation of non-cognate genes, the PFLU1132 transcription factor will need to first saturate its native binding sites before it can engage in promiscuous interactions with FleQ-controlled genes. High expression of active transcription factor can provide these conditions, elevating the concentration of transcription factor in the cell to permit more frequent promiscuous action[47,48].

Whilst many of the 2nd step mutations detailed above may impact PFLU1131/2 expression indirectly (RpoC, Rho), the PFLU1131/2 promoter mutation directly effects expression of the two component system. This mutation is a 1.59Kbp deletion resulting in total loss of the PFLU1130 gene, 227bp of the 3' end of PFLU1129,183bp of the 5' end of PFLU1131 and loss of 119bp the intergenic region between the PFLU1127/8/9 and PFLU1130/1/2 operons (Fig.5A). Whilst creating at PFLU1129/PFLU1131 partial fusion protein, the rest of the PFLU1131 ORF is preserved, starting with a GTG codon and with no impact on the coding sequence of the three catalytic domains. This deletion results in PFLU1131/2 becoming part of the PFLU1127/8/9 operon. Upstream of this new combined operon sits a predicted RpoN binding site[43], so that this deletion results in PFLU1132 being positioned in an operon under the control of RpoN and an unknown RpoN-EBP.

The deleted intergenic region is predicted to contain a rho-dependent terminator (Supplementary Excel file E3) which may typically prevent transcriptional readthrough from PFLU1127/8/9 into PFLU1130/1/2. The *rho* mutation observed in another 2nd step motile strain may also act to increase readthrough. The PFLU1127/8/9 operon is upregulated upon the 1st step *PFLU1131*-del15 mutation, as are other nearby operons predicted to be controlled by RpoN[43] (PFLU1122/3/4 and PFLU1125/6 – Fig.5B), possibly due to promiscuous activation due to the hyperactive PFLU1132 transcription factor. The PFLU1130/1/2 operon is also upregulated by the *PFLU1131*-del15 mutation, despite lacking a predicted RpoN binding site. This increase in PFLU1130/1/2 expression may be due to increased readthrough from increase PFLU1127/8/9 expression in the 1st step mutant.

The 1.59Kbp 2nd step mutation to the PFLU1131/2 promoter results in significant upregulation of both PFLU1131 and PFLU1132 with 4.26 and 3.79 Log2Fold increases in expression compared to AR2 respectively (Fig.5B). This upregulation of the PFLU1131/2 two-component system corresponds with the 60% increase in expression RpoN-EBP controlled genes discussed above (Fig.4E), indicating an increase in promiscuous transcription factor activity. The PFLU1583 Δ48-74 mutation does not significantly impact PFLU1131/2 expression (Fig.5B), so this mutation likely influences the promiscuity of the PFLU1132 transcription factor through a separate mechanism.

Increasing transcription factor expression in motile mutants results in faster motility phenotypes

To test the importance of transcription factor gene expression in gain of promiscuous activity, we made use of the complementation strains presented in Fig.1C and supplementary Fig.S1. These experiments produced strains where the genes *ntrC* and *PFLU1132* are deleted from their native loci and reintroduced on an L-rhamnose titratable promoter system. This was done in the presence of their respective kinase mutations, *ntrB*-T97P and *PFLU1131*-del15. Concentration of L-rhamnose added to the media positively correlated with distance moved (Fig.6A) for both the NtrC expression system ($\rho = 0.984$, Spearman's test) and the PFLU1132 expression system ($\rho = 0.865$, Spearman's test). To ensure increasing L-rhamnose concentration results in increasing expression from the RhaSR-PrhaBAD system in our bacterial strain backgrounds, a control was run testing LacZ activity with increasing L-rhamnose concentration for a miniTn7[*rhaSR-PrhaBAD-stRBS-lacZ*] construct in AR2 and AR2ΔntrC backgrounds. In both cases LacZ activity positively correlated with L-rhamnose concentration ($\rho = 0.981$ and $\rho = 0.955$ respectively, Spearman's tests), indicating that increasing L-rhamnose concentration results in increasing expression of the gene introduced in the *rhaSR-PrhaBAD* expression system (Fig.6C). This allows us to conclude that increasing expression of NtrC and PFLU1132 in the presence of their 1st step kinase mutations granted stronger motility phenotypes, likely through stronger promiscuous action.

Discussion

Essential to understanding why some organisms adapt faster than others in the face of environmental challenges is knowledge of how GRNs and regulatory systems evolve. Evolutionary innovation of transcription factors within these networks depends on gain of functional promiscuity as a first step, however the rules of when and where such properties develop within the context of the GRN is currently unknown. Using our microbial model system, we have identified two available pathways for rescuing flagellar motility in a maladapted Δ*fleQ* network; through mutations to the NtrBC and PFLU1131/2 two component systems. However, one pathway – the NtrBC pathway – is always used to the exclusion of the other when both are present. By comparing these available evolutionary pathways, we can resolve why NtrC is more accessible for gaining promiscuity and rewiring due to its position within its GRN. In addition, we can identify three key properties that facilitate transcription factor promiscuity and opportunities for evolutionary innovation.

Firstly, our results confirm the importance of shared structural homology between the absent and recruited transcription factor. In our system, the shared structural homology of NtrC and PFLU1132 with FleQ grants a basal affinity for regulating flagellar genes that evolution can act upon. Whilst structural similarity to FleQ is clearly important, it is not the deciding factor for promiscuity in our model as there are other RpoN-EBPs with closer homology to FleQ in *P. fluorescens*[37]. It is possible that the RpoN-EBPs represent a hyper-evolvable family of transcription factors through the nature of their connectivity and mechanism of transcriptional regulation via RpoN. To investigate this evolutionary rescue of a non-RpoN dependent trait would need to be tested using a different maladapted GRN rescue model. **Secondly, hyperactivation of the co-opted transcription factor is required for unlocking promiscuity**. In our model system, first step mutations that granted a slow motile phenotype were always achieved through hyperactivation of a cognate kinase. Most transcription factors are subject to some form of post-translational control, most commonly through phosphorylation of a REC receiver domain in bacteria[49]. Promiscuous binding may occur transiently or only briefly, so ensuring that the transcription factor is constantly activated increases the likelihood of these non-cognate regulatory events leading to altered gene expression. **Finally - and perhaps most importantly - high gene expression of the transcription factor is needed for promiscuity**. High expression of a transcription factor can lead to saturation of native binding sites. Any unbound, active transcription factor is then free to engage in promiscuous activity. These low affinity interactions will occur more readily if there is a higher concentration of transcription factor present, following laws of mass action kinetics[50]. Typically, transcription factors are expressed at low levels in bacteria[51], due to only a small protein copy number being needed

for a small number of binding sites in most cases[52]. Transcription factor expression level may well be optimised at a low level to prevent mis-regulation. Indeed, highly expressed and connected transcription factors were found to result in greater network perturbations when synthetically rewired[15], which may reflect a greater ability to act promiscuously. Many molecular biology studies use overexpression/overactivation of proteins of interest as a means to determining their biological functions[53]. Our findings also suggest that this approach should perhaps be used more cautiously, certainly when studying transcription factor function.

Aside from mutations directly targeting activation and expression of PFLU1131/2, we observed several distinct strategies that enhance PFLU1132 promiscuity through global regulator changes. These strategies likely impact the wider GRN and transcriptome significantly, with mutations to RpoC, RpsK and Rho being particularly striking. RpoC and RpsK mutations may act to increase expression of most genes including PFLU1131/2 and the flagellar genes. Mutations to the PFLU1583/4 anti-sigma factor system also indicated a strategy to boost promiscuity that would have a significant impact of the wider GRN. Global regulators are often sites of mutation and innovation[54,55], with mutations affecting sigma factor function in particular having large impacts on global gene expression which can drive phenotypic innovation[56]. Changes to sigma factors have been shown to facilitate GRN remodelling, as was the case for a study of *P. aeruginosa* during adaptation to the CF lung[13]. Our study highlights for the first time that such mutational strategies can also achieve increases in promiscuity for specific transcription factors to facilitate evolutionary innovation.

Each of these mutational strategies that enhanced functional promiscuity in PFLU1132 acted to alter the pre-existing GRN architecture around the transcription factor, whether at a local level by altering the PFLU1131/2 promoter or at a global regulatory level. This highlights that the pre-existing GRN is a key constraining factor for gain of the properties needed for functional promiscuity and therefore for evolutionary innovation of a transcription factor. As environmental signals are transduced through GRNs, the pre-existing network structure will also determine which transcription factors may become active and promiscuous under a given set of conditions, acting to reveal particular network components to selection[28,57]. Transcription factors are often subject to a form of positive or negative autoregulation[58]. Achieving high expression will therefore be far easier for those that are under positive feedback control. Typically this autoregulation is prevented from engaging in runaway feedback by a negative repressor[59–61], including in the case of NtrBC. Loss of repression will lead to runaway feedback and high expression, whereas to achieve a similar effect in a negative autoregulator, precise promoter mutations[62] will be needed and a similarly high expression level will be much harder to achieve. In this manner, pre-existing GRN motifs can

be predicting to significantly bias the evolvability of the transcription factors under their control. We observe this bias in our model system. The more evolvable pathway in gaining promiscuity is the NtrBC system in which the negative repressor GlnK prevents runaway *ntrBC* autoregulation by inhibiting NtrB[63]. All previously observed NtrBC pathway mutations were single *de-novo* mutations that act to remove repression through GlnK, both directly or indirectly[34,38,39]. These have a dual effect, both hyperactivating the NtrB kinase, and strongly upregulating *ntrBC* expression through their positive autoregulatory loop. Mutational loss of a negative repressor leading to overactivity and overexpression of a positively autoregulated transcription factor has also been observed to drive adaptation in several experimental and clinical settings[35,64,65]. In contrast, we have identified that to achieve the same effect of high activity and high expression in PFLU1131/2, multiple mutations were required. This highlights the clear advantage NtrBC has in becoming promiscuous compared to PFLU1131/2.

Our study provides empirical evidence for mechanisms and evolutionary drivers of regulatory evolution. GRNs are key to the adaptability of organisms, and our results reveal the genetic regulatory factors that generate evolvable control systems. We find that feedback loops that can generate high expression of regulators after a small number of mutations can facilitate rapid innovation and adaptation. Previously, we have hypothesised that some transcription factors may be ‘primed’ for evolutionary innovation by virtue of the GRN architecture they exist within[28]. Our data support this by suggesting that promiscuity can be more easily evolved in transcription factors with positive feedback structures. These findings can also be utilised in the design of genetic regulatory systems to have reduced evolvability, a key future goal for synthetic biology projects where the evolutionary change of a constructed system is detrimental[66,67]. Designing systems that avoid easily accessible routes to hyperactivation and hyperexpression of their components could help prevent rewiring of engineered GRNs. Pre-existing GRN structure can be expected to significantly bias patterns of diversification in regulatory systems, and subsequently shape genomic evolution by influencing the speed in which genetic circuitry can adapt to changing conditions. This is highly important for populations adapting during sudden changes such as niche transitions including the emergence of new pathogens. Our data moves us closer to the eventual goal of empirically building a set of principles by which GRNs adapt and offers unique insights into how these systems can function and influence the evolution of their components.

Materials and Methods

Strains and culture conditions

All ancestral lines in this study are derived from *Pseudomonas fluorescens* AR2 (SBW25Δ*fleQ* IS-ΩKm-hah: PFLU2552). AR2 lacks the flagellum master regulator FleQ and possesses a transposon-insertional disruption of the gene *viscB* (PFLU2552), rendering it immotile as detailed previously[34,68]. All strains were cultured on lysogeny broth (LB; Miller) media at 27°C. *Escherichia coli* strains for cloning were cultured on LB media at 37°C. Strains, plasmids and media supplements are detailed in Supplementary Excel file E4.

Knockout mutant construction by two-step allelic exchange

Knockouts of ntrC and PFLU1132 were achieved using two-step allelic exchange following the protocol of Hmelo et al., 2015 with some alterations. In brief, 400bp flanking regions up- and down-stream of the target gene were amplified and joined by strand overlap extension (SOE) PCR to create a knockout allele. This was inserted into the allelic-exchange suicide vector pTS1 (Containing *sacB* and *tetR*) by SOE-cloning[70] to create the knockout plasmid, and transformed into the conjugal *E. coli* strain ST18 by chemical-competence heat-shock. Knockout plasmids were transferred from *E. coli* ST18 to AR2 by two-parent puddle-mating and merodiploids selected for on LB supplemented with kanamycin sulphate and tetracycline hydrochloride but lacking 5-ALA supplement required for growth of the *E. coli* ST18 auxotrophic Δ*hemA* mutant. Merodiploids were cultured overnight in LB broth lacking tetracycline selection and diluted before spread plating onto NSLB media supplemented with 15% w/v sucrose for *sacB* levansucrase-mediated counterselection of the pTS1 plasmid backbone. Sucrose-resistant colonies were isolated and screened for tetracycline sensitivity. Chromosomal presence of the knockout allele and absence of target gene coding sequences was confirmed by colony PCR. Knockout of ntrC left an ORF encoding an 11 amino acid truncated protein, and knockout of PFLU1132 left an ORF encoding an 8 amino acid truncated protein. All other genomic features including operon structure and terminator regions were left intact.

Motility evolution experiments

AR2ΔntrC and AR2ΔntrCΔPFLU1132 were challenged to rescue motility in the absence of the FleQ master flagellar regulator on soft agar, as described previously[34,38]. Pure colonies were picked and inoculated into 0.25% agar LB plates made as described in Alsohim et al., 2014, and incubated at 27°C. Plates were checked a minimum of twice daily for motility, recording time to emergence. Motile zones were sampled immediately and always from the leading edge. Motile isolates were streaked on LB agar, and a pure colony picked and stored

at -80°C as glycerol stocks of LB overnight cultures. All subsequent analysis was conducted on these pure motile isolates. Experiment was run for six weeks and any replicates without motility after this cut-off recorded as having not evolved.

Bacterial growth and motility fitness assays

The fitness of the motility phenotype of evolved AR2ΔntrC strains was tested by measuring distance moved over 24 hrs of incubation in 0.25% agar LB plates. Six biological replicates of each strain were grown as separate overnight cultures. Cultures were adjusted to an OD595 = 1 and resuspended in PBS. Soft agar plates were inoculated with 1 µL of these suspensions, by piercing the surface of the plate with the pipette tip, and then effusing the sample into the gap left by the tip. Plates were incubated for 24 h at 27°C, and photographs taken of motile zones. Surface area moved was then calculated from the radius of the concentric motile zone measured from these images ($A = \pi r^2$). Values were square root transformed before plotting.

Growth in shaking LB broth was measured by inoculating 99 µL of sterile LB broth with 1 µL of the OD595=1 PBS cell suspensions for each replicate in a 96-well plate. Plates were incubated at 27°C with 180 rpm shaking in a plate reader, recording OD595 every ten minutes. Area under the bacterial growth curve was calculated using the growthcurver package in R and plotted[71]. Area under the bacterial growth curve provides a metric for fitness, as it accounts for the characteristics of lag- and log-phase growth, as well as the final carrying capacity of the population.

Mutation identification by whole genome resequencing and PCR amplicon sanger sequencing

To identify motility rescuing mutations, genomic DNA was extracted from motile strains and their ancestral strain using the Thermo Scientific GeneJET Genomic DNA Purification Kit. Genomic DNA was quality checked using BR dsDNA Qubit spectrophotometry to determine concentration and nanodrop spectrophotometry to determine purity. Illumina HiSeq sequencing was provided by MicrobesNG (Birmingham, UK), with a minimum 30x coverage and a sample average of 114x coverage. Returned paired-end reads were aligned to the *P. fluorescens* SBW25 reference genome[44] using the Galaxy platform[72]. INDELs were identified using the integrated genomics viewer[73]. For SNP identification, the variant calling software SNIPPY was used with default parameters (Seemann, 2015). Protein domains affected by mutations detailed in this article were predicted from amino acid sequences using pfam, SMART and BLASTp. For a subset of additional motile mutants, PCR amplification and subsequent sanger sequencing of PFLU1131 was performed as this gene was mutated in all

previously sequenced strains. This was done using the service provided by Eurofins Genomics. PCR amplicons were purified using the Monarch® PCR & DNA Cleanup Kit (New England Biolabs). Mutations were identified by alignment of the returned PFLU1131 sequence against the *P. fluorescens* SBW25 reference genome using NCBI BLAST.

Three 2nd step isolates derived from line 24 were sampled due to observed colony morphology variation. After whole genome re-sequencing it became apparent that this variation was due to the presence of a *mutS* mutation and not due to diverse motility strategies. These isolates all have the mutation PFLU1132-D111G (Supplementary excel file E1, isolate ID's 24-S1, 24-LCV, and 24-SCV), however this is likely not due to parallel evolution, but instead a single instance of a PFLU1132-D111G mutant that then diversified due to *mutS* mutation. For all subsequent analysis these three isolates were treated as a single instance of PFLU1132-D111G.

RNA-sequencing

Whole-cell RNA was extracted from 20 OD units of *P. fluorescens* cultures in mid-log phase growth (OD_{595} ~1.5). Cultures were incubated in LB broth at 27°C and 180 rpm shaking. Extractions were performed for biological triplicates of each strain of interest. Upon reaching the desired OD, growth and RNA expression was halted by addition of a ½ culture volume of ice-cold killing buffer (20 mM NaN_3, 20 mM Tris-HCl, 5 mM $MgCl_2$). Cells were pelleted, and the killing buffer removed. A lysis buffer of β-mercaptoethanol in buffer RLT from the Qiagen RNeasy extraction kit was used to resuspend pellets, which were then lysed by bead-milling at 4500 rpm for 45 s with lysing matrix B. Lysates were spun through columns from the Qiagen RNeasy extraction kit, and the extraction completed following the RNeasy kit protocol. A DNase I treatment step was included between washes with buffer RW1, by adding DNase I directly to the column from the RNase-free DNase kit (Qiagen) following kit protocol. Samples were eluted in nuclease-free water, and subsequently treated with TURBO DNase from the Turbo DNA-free kit (Invitrogen) following kit protocols.

Purified RNA concentration was measured by Qubit RNA BR assay (Thermo-scientific), RNA quality by nanodrop spectrophotometry, and RNA integrity by agarose gel electrophoresis. RNA sequencing was provided by Oxford Genomics (Oxford UK). Samples were ribodepleted, and the mRNA fraction converted to cDNA with dUTP incorporated during second strand synthesis. The cDNA was end-repaired, A-tailed and adapter-ligated. Prior to amplification, samples underwent uridine digestion. The prepared libraries were size selected, multiplexed and QC'ed before paired end sequencing over one unit of a flow cell in a NovaSeq 6000. Returned data was quality checked before it was returned in fastq format. Reference-based transcript assembly was performed using the Rockhopper RNA

sequencing analysis software[74]. Differential gene expression analysis was performed using the DESeq2 package of the R statistical coding language to create lists of differentially expressed genes and corresponding Log2Fold changes in gene expression.

Rhamnose-inducible expression system constructs

Complementation of *ntrC* and *PFLU1132* knockouts, as well as inducible expression experiments were performed by single-copy chromosomal re-introduction of the RpoN-EBP genes under control of a *rhaSR-PrhaBAD* L-rhamnose inducible expression system[75]. The RpoN-EBP ORF was amplified and a strong ribosome binding site (stRBS) introduced upstream after a 7bp short spacer between the site and the start codon. This construct was placed downstream of the *PrhaBAD* promoter on the miniTn7 suicide-vector pJM220 (obtained from the Addgene plasmid repository, plasmid #110559) by restriction-ligation and transferred to *E. coli* DH5α by chemical-competence heat-shock. The miniTn7 transposon containing the *rhaSR* genes and the PrhaBAD-stRBS-RpoN-EBP construct were transferred to the *P. fluorescens* AR2 chromosome by transposonal insertion downstream of the *glmS* gene via four-parent puddle-mating conjugation[76]. The relevant *E. coli* DH5α pJM220-derived plasmid donor was combined with recipient *P. fluorescens* AR2 strains, transposition helper *E. coli* SM10 λpir pTNS2 and conjugation helper *E. coli* SP50 pRK2073, and Gentamicin resistant *Pseudomonas* selected for on LB supplemented with Gentamicin sulphate and Kanamycin sulphate. Chromosomal insertion of the correct miniTn7 transposon was confirmed by colony PCR.

β-galactosidase activity assay

To validate the expression dynamics of the rhaSR-PrhaBAD system in the genetic backgrounds used in this work, the same system described above was introduced into *P. fluorescens* AR2 and AR2ΔntrC with lacZ under the control of PrhaBAD. This was obtained from the pJM230 plasmid[75] (obtained from the Addgene plasmid repository, plasmid #110560). Overnight cultures of each strain were set up in biological triplicate, and used to inoculate 9 mL cultures of LB broth supplemented with a range of L-rhamnose concentrations at a bacterial OD_{600} = 0.05. Cultures were incubated shaking 180 r.p.m. at 27°C until they reached and OD_{600} = 0.5, 2 mL was spun down and frozen at -20°C until needed. Cell pellets were resuspended in Z-buffer (60mM Na_2HPO_4*$7H_2O$, 40mM NaH_2PO4*H_2O, 10mM KCl, 1mM $MgSO_4$*$7H_2O$), OD_{600} recorded, and 400µL cells added to 600µL fresh Z-buffer. Cell suspension was lysed with addition of 40 µL Chloroform, 20 µL 0.1% SDS and vortexing followed by 10 minutes incubation at 30°C. β-galactosidase activity assay was begun with addition of 200µL o-nitrophenyl-β-D-galactopyraniside (ONPG) and

incubated at 30°C for 20 minutes. If reactions become visibly yellow they were halted immediately by addition of 500 µL Na_2CO_3 and the time recorded. All other reactions were halted in this manner after 20 minutes. Reactions were centrifuged to remove debris and 900 µL added to a cuvette and A_{420} measured. LacZ reporter activity (MU) was calculated by: MU = (A_{420} * 1000) / (Time(mins) * 0.4 * OD_{600} of initial Z-buffer cell suspension).

Statistical analyses and data handling

All statistical analysis and data handling was performed using R core statistical packages and the Dunn.test package. Shapiro-Wilks normality tests were performed to confirm non-normality of datasets. To compare group medians for more than two groups, a Kruskal-Wallis test with post-hoc Dunn test and Benjamini-Hochberg correction was performed, with a $P \leq 0.025$ taken to indicate significance. For comparisons of two groups only, a two-sample Wilcox test was used.

For analysis of transcriptomic data, RNA sequencing output was run through Rockhopper for reference-based transcript assembly. Transcript count data were then analysed using DESeq2 package in R, for differential gene expression analysis, as well as generation of PCA and Volcano plots. A gene was taken as being differentially expressed if it differed by a Log2Fold change of ≥1 or ≤ -1 between two samples (representing a doubling or halving of expression level respectively) with an adjusted P-value of ≤0.001.

To identify whether Rho dependent terminators were present downstream of genes of interest, the RhoTermPredict algorithm[77] was run on the *P. fluorescens* SBW25 reference genome in Python.

<u>References</u>

1. Sheppard SK, Guttman DS, Fitzgerald JR. Population genomics of bacterial host adaptation. *Nat Rev Genet*. 2018;19(9):549-565. doi:10.1038/s41576-018-0032-z

2. Cohan FM, Koeppel AF. The Origins of Ecological Diversity in Prokaryotes. *Curr Biol*. 2008;18(21):1024-1034. doi:10.1016/j.cub.2008.09.014

3. Martin CH, Richards EJ. The Paradox behind the Pattern of Rapid Adaptive Radiation: How Can the Speciation Process Sustain Itself Through an Early Burst? *Annu Rev Ecol Evol Syst*. 2019;50:569-593. doi:10.1146/annurev-ecolsys-110617-062443

4. Fang X, Sastry A, Mih N, et al. Global transcriptional regulatory network for Escherichia coli robustly connects gene expression to transcription factor activities. *Proc Natl Acad Sci U S A*. 2017;114(38):10286-10291. doi:10.1073/pnas.1702581114

5. López-Maury L, Marguerat S, Bähler J. Tuning gene expression to changing environments: From rapid responses to evolutionary adaptation. *Nat Rev Genet*. 2009;10(1):68. doi:10.1038/nrg2500

6. Martínez-Antonio A, Janga SC, Salgado H, Collado-Vides J. Internal-sensing machinery directs the activity of the regulatory network in Escherichia coli. *Trends Microbiol*. 2006;14(1):22-27. doi:10.1016/j.tim.2005.11.002

7. Mcdonald JMC, Reed RD. Patterns of selection across gene regulatory networks. *Semin Cell Dev Biol*. 2022;(xxxx). doi:10.1016/j.semcdb.2022.03.029

8. Isalan M, Lemerle C, Michalodimitrakis K, et al. Evolvability and hierarchy in rewired bacterial gene networks. *Nature*. 2008;452(7189):840-845. doi:10.1038/nature06847

9. Mack KL, Ballinger MA, Phifer-Rixey M, Nachman MW. Gene regulation underlies environmental adaptation in house mice. *Genome Res*. 2018;28(11):1636-1645. doi:10.1101/gr.238998.118

10. Luo X, Song R, Acar M. Multi-component gene network design as a survival strategy in diverse environments. *BMC Syst Biol*. 2018;12(1):85. doi:10.1186/s12918-018-0609-3

11. Ali F, Seshasayee ASN. Dynamics of genetic variation in transcription factors and its implications for the evolution of regulatory networks in Bacteria. *Nucleic Acids Res*. 2021;48(8):4100-4114. doi:10.1093/NAR/GKAA162

12. Turkarslan S, Reiss DJ, Gibbins G, et al. Niche adaptation by expansion and reprogramming of general transcription factors. *Mol Syst Biol*. 2011;7(554):1-16. doi:10.1038/msb.2011.87

13. Damkiær S, Yang L, Molin S, Jelsbak L. Evolutionary remodeling of global regulatory networks during long-term bacterial adaptation to human hosts. *Proc Natl Acad Sci U S A*. 2013;110(19):7766-7771. doi:10.1073/pnas.1221466110

14. Johnson AD. The rewiring of transcription circuits in evolution. *Curr Opin Genet Dev*. 2017;47:121-127. doi:10.1016/j.gde.2017.09.004

15. Baumstark R, Hänzelmann S, Tsuru S, et al. The propagation of perturbations in rewired bacterial gene networks. *Nat Commun*. 2015;6:1-5. doi:10.1038/ncomms10105

16. Kruzel EK, Giles SS, Hull CM. Analysis of cryptococcus neoformans sexual development reveals rewiring of the pheromone-response network by a change in

transcription factor identity. *Genetics*. 2012;191(2):435-449. doi:10.1534/genetics.112.138958

17. Reece-Hoyes JS, Pons C, Diallo A, et al. Extensive rewiring and complex evolutionary dynamics in a C.elegans multiparameter transcription factor network. *Mol Cell*. 2013;51(1):116-127. doi:10.1016/j.molcel.2013.05.018

18. Tebung WA, Choudhury BI, Tebbji F, Morschhäuser J. Rewiring of the Ppr1 Zinc Cluster Transcription Factor from Purine Catabolism to Pyrimidine Biogenesis in the Saccharomycetaceae. *Curr Biol*. 2016;26(13):1677-1687. doi:10.1016/j.cub.2016.04.064

19. Adhikari S, Erill I, Curtis PD. Transcriptional rewiring of the GcrA/CcrM bacterial epigenetic regulatory system in closely related bacteria. *PLoS Genet*. 2021;17(3):1-30. doi:10.1371/journal.pgen.1009433

20. Martchenko M, Levitin A, Hogues H, Nantel A, Whiteway M. Transcriptional Rewiring of Fungal Galactose-Metabolism Circuitry. *Curr Biol*. 2007;17(12):1007-1013. doi:10.1016/j.cub.2007.05.017

21. Hsu PC, Lu TC, Hung PH, et al. Plastic Rewiring of Sef1 Transcriptional Networks and the Potential of Nonfunctional Transcription Factor Binding in Facilitating Adaptive Evolution. *Mol Biol Evol*. 2021;38(11):4732-4747. doi:10.1093/molbev/msab192

22. Voordeckers K, Pougach K, Verstrepen KJ. How do regulatory networks evolve and expand throughout evolution? *Curr Opin Biotechnol*. 2015;34:180-188. doi:10.1016/j.copbio.2015.02.001

23. Copley SD. An evolutionary biochemist's perspective on promiscuity. *Trends Biochem Sci*. 2015;40(2):72-78. doi:10.1016/j.tibs.2014.12.004

24. Guzmán GI, Sandberg TE, LaCroix RA, et al. Enzyme promiscuity shapes adaptation to novel growth substrates. *Mol Syst Biol*. 2019;15(4):1-14. doi:10.15252/msb.20188462

25. Shimada T, Ishihama A, Busby SJW, Grainger DC. The Escherichia coli RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Res*. 2008;36(12):3950-3955. doi:10.1093/nar/gkn339

26. Copley SD. The physical basis and practical consequences of biological promiscuity. *Phys Biol*. 2020;17(5). doi:10.1088/1478-3975/ab8697

27. Aharoni A, Gaidukov L, Khersonsky O, Gould SM, Roodveldt C, Tawfik DS. The

"evolvability" of promiscuous protein functions. *Nat Genet*. 2005;37(1):73-76. doi:10.1038/ng1482

28. Taylor TB, Shepherd MJ, Jackson RW, Silby MW. Natural selection on crosstalk between gene regulatory networks facilitates bacterial adaptation to novel environments. *Curr Opin Microbiol*. 2022;67:102140. doi:10.1016/j.mib.2022.02.002

29. Capra EJ, Perchuk BS, Skerker JM, Laub MT. Adaptive mutations that prevent crosstalk enable the expansion of paralogous signaling protein families. *Cell*. 2012;150(1):222-232. doi:10.1016/j.cell.2012.05.033

30. Zarrinpar A, Park SH, Lim WA. Optimization of specificity in a cellular protein interaction network by negative selection. *Nature*. 2003;426(6967):676-680. doi:10.1038/nature02178

31. Pougach K, Voet A, Kondrashov FA, et al. Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. *Nat Commun*. 2014;5:1-11. doi:10.1038/ncomms5868

32. Brunkard JO, Runkel AM, Zambryski PC. A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science (80-)*. 2015;347(6222):645-649. doi:10.1126/science.1255437

33. Alhindi T, Zhang Z, Ruelens P, et al. Protein interaction evolution from promiscuity to specificity with reduced flexibility in an increasingly complex network. *Sci Rep*. 2017;7(February):1-15. doi:10.1038/srep44948

34. Taylor TB, Mulley G, Dills AH, et al. Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system. *Science (80-)*. 2015;347(6225):1014-1017. doi:10.1126/science.1259145

35. Bretl DJ, Ladd KM, Atkinson SN, Müller S, Kirby JR. Suppressor mutations reveal an NtrC-like response regulator, NmpR, for modulation of Type-IV Pili-dependent motility in Myxococcus xanthus. *PLoS Genet*. 2018;14(10):1-28. doi:10.1371/journal.pgen.1007714

36. Naren N, Zhang XX. Role of a local transcription factor in governing cellular carbon/nitrogen homeostasis in Pseudomonas fluorescens. *Nucleic Acids Res*. 2021;49(6):3204-3216. doi:10.1093/nar/gkab091

37. Taylor T, Mulley G, McGuffin L, et al. Evolutionary rewiring of bacterial regulatory networks. *Microb Cell*. 2015;2(7):256-258. doi:10.15698/mic2015.07.215

38. Horton JS, Flanagan LM, Jackson RW, Priest NK, Taylor TB. A mutational hotspot that determines highly repeatable evolution can be built and broken by silent genetic changes. *Nat Commun*. 2021;12(1). doi:10.1038/s41467-021-26286-9

39. Shepherd MJ, Horton JS, Taylor TB. A near-deterministic mutational hotspot in Pseudomonas fluorescens is constructed by multiple interacting genomic features. *Mol Biol Evol*. 2022;39(6):1-7. doi:10.1093/molbev/msac132

40. Igler C, Lagator M, Tkačik G, Bollback JP, Guet CC. Evolutionary potential of transcription factors for gene regulatory rewiring. *Nat Ecol Evol*. 2018;2(10):1633-1643. doi:10.1038/s41559-018-0651-y

41. Lind PA, Farr AD, Rainey PB. Experimental evolution reveals hidden diversity in evolutionary pathways. *Elife*. 2015;4:1-17. doi:10.7554/elife.07074

42. Kim DJ, Forst S. Genomic analysis of the histidine kinase family in bacteria and archaea. *Microbiology*. 2001;147(5):1197-1212. doi:10.1099/00221287-147-5-1197

43. Jones J, Studholme DJ, Knight CG, Preston GM. Integrated bioinformatic and phenotypic analysis of RpoN-dependent traits in the plant growth-promoting bacterium Pseudomonas fluorescens SBW25. *Environ Microbiol*. 2007;9(12):3046-3064. doi:10.1111/j.1462-2920.2007.01416.x

44. Silby MW, Cerdeño-Tárraga AM, Vernikos GS, et al. Genomic and genetic analyses of diversity and plant interactions of Pseudomonas fluorescens. *Genome Biol*. 2009;10(5):1-16. doi:10.1186/gb-2009-10-5-r51

45. Oliver A, Baquero F, Blázquez J. The mismatch repair system (mutS, mutL and uvrD genes) in Pseudomonas aeruginosa: Molecular characterization of naturally occurring mutants. *Mol Microbiol*. 2002;43(6):1641-1650. doi:10.1046/j.1365-2958.2002.02855.x

46. Luján AM, Maciá MD, Yang L, Molin S, Oliver A, Smania AM. Evolution and adaptation in pseudomonas aeruginosa biofilms driven by mismatch repair system-deficient mutators. *PLoS One*. 2011;6(11). doi:10.1371/journal.pone.0027842

47. Gutierrez PS, Monteoliva D, Diambra L. Cooperative Binding of Transcription Factors Promotes Bimodal Gene Expression Response. *PLoS One*. 2012;7(9). doi:10.1371/journal.pone.0044812

48. Kribelbauer JF, Rastogi C, Bussemaker HJ, Mann RS. Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. *Annu Rev Cell Dev Biol*.

2019;35:357-379. doi:10.1146/annurev-cellbio-100617-062719

49. Galperin MY. Structural classification of bacterial response regulators: Diversity of output domains and domain combinations. *J Bacteriol*. 2006;188(12):4169-4182. doi:10.1128/JB.01887-05

50. Kim JK, Tyson JJ. Misuse of the Michaelis–Menten rate law for protein interaction networks and its remedy. *PLoS Comput Biol*. 2020;16(10):1-21. doi:10.1371/journal.pcbi.1008258

51. Ishihama Y, Schmidt T, Rappsilber J, et al. Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics*. 2008;9:1-17. doi:10.1186/1471-2164-9-102

52. Alkema WBL, Lenhard B, Wasserman WW. Regulog analysis: Detection of conserved regulatory networks across bacteria: Application to Staphylococcus aureus. *Genome Res*. 2004;14(7):1362-1373. doi:10.1101/gr.2242604

53. Prelich G. Gene overexpression: Uses, mechanisms, and interpretation. *Genetics*. 2012;190(3):841-854. doi:10.1534/genetics.111.136911

54. Philippe N, Crozat E, Lenski RE, Schneider D. Evolution of global regulatory networks during a long-term experiment with Escherichia coli. *BioEssays*. 2007;29(9):846-860. doi:10.1002/bies.20629

55. Harrison E, Guymer D, Spiers AJ, Paterson S, Brockhurst MA. Parallel Compensatory Evolution Stabilizes Plasmids across the Parasitism-Mutualism Continuum. *Curr Biol*. 2015;25(15):2034-2039. doi:10.1016/j.cub.2015.06.024

56. Pletnev P, Pupov D, Pshanichnaya L, et al. Rewiring of growth-dependent transcription regulation by a point mutation in region 1.1 of the housekeeping σ factor. *Nucleic Acids Res*. 2020;48(19):10802-10819. doi:10.1093/nar/gkaa798

57. Jothi R, Balaji S, Wuster A, et al. Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. *Mol Syst Biol*. 2009;5(294). doi:10.1038/msb.2009.52

58. Macneil LT, Walhout AJM. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res*. 2011;21:645-657. doi:10.1101/gr.097378.109.21

59. Rao SD, Igoshin OA. Overlaid positive and negative feedback loops shape dynamical properties of PhoPQ two-component system. *PLoS Comput Biol*. 2021;17(1 December):1-18. doi:10.1371/journal.pcbi.1008130

60. Weyder M, Prudhomme M, Bergé M, Polard P, Fichant G. Dynamic modeling of Streptococcus pneumoniae competence provides regulatory mechanistic insights into its tight temporal regulation. *Front Microbiol*. 2018;9(JUL):1-25. doi:10.3389/fmicb.2018.01637

61. diCenzo GC, Sharthiya H, Nanda A, Zamani M, Finan TM. PhoU allows rapid adaptation to high phosphate concentrations by modulating PstSCAB transport rate in Sinorhizobium meliloti. *J Bacteriol*. 2017;199(18):1-20. doi:10.1128/JB.00143-17

62. Lamrabet O, Plumbridge J, Martin M, Lenski RE, Schneider D, Hindre T. Plasticity of promoter-core sequences allows bacteria to compensate for the loss of a key global regulatory gene. *Mol Biol Evol*. 2019;36(6):1121-1133. doi:10.1093/molbev/msz042

63. Hervás AB, Canosa I, Santero E. Regulation of glutamate dehydrogenase expression in Pseudomonas putida results from its direct repression by NtrC under nitrogen-limiting conditions. *Mol Microbiol*. 2010;78(2):305-319. doi:10.1111/j.1365-2958.2010.07329.x

64. Patel V, Matange N. Adaptation and compensation in a bacterial gene regulatory network evolving under antibiotic selection. *Elife*. 2021;10:1-27. doi:10.7554/eLife.70931

65. Schofield MC, Rodriguez DQ, Kidman AA, et al. The anti-sigma factor MucA is required for viability in Pseudomonas aeruginosa. *Mol Microbiol*. 2021;116(2):550-563. doi:10.1111/mmi.14732

66. Gallup O, Ming H, Ellis T. Ten future challenges for synthetic biology. *Eng Biol*. 2021;5(3):51-59. doi:10.1049/enb2.12011

67. Chalovich JM, Eisenberg E. Engineering reduced evolutionary potential for synthetic biology. *Biophys Chem*. 2005;257(5):2432-2437. doi:10.1039/c3mb70606k.Engineering

68. Alsohim AS, Taylor TB, Barrett GA, et al. The biosurfactant viscosin produced by Pseudomonas fluorescensSBW25 aids spreading motility and plant growth promotion. *Environ Microbiol*. 2014;16(7):2267-2281. doi:10.1111/1462-2920.12469

69. Hmelo LR, Borlee BR, Almblad H, et al. Precision-engineering the Pseudomonas aeruginosa genome with two-step allelic exchange. *Nat Protoc*. 2015;10(11):1820-1841. doi:10.1038/nprot.2015.115

70. Bryksin A V., Matsumura I. Overlap extension PCR cloning: a simple and reliable way

to create recombinant plasmids. *Biotechniques*. 2010;48(6):463-465. doi:10.2144/000113418.Overlap

71. Sprouffske K, Wagner A. Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*. 2016;17(1):17-20. doi:10.1186/s12859-016-1016-7

72. Afgan E, Baker D, Batut B, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res*. 2018;46(W1):W537-W544. doi:10.1093/nar/gky379

73. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26. doi:10.1038/nbt.1754

74. Brian Tjaden. A Computational System for Identifying Operons Based on RNA- Seq Data. *Methods*. 2020;176:60-72. doi:10.1016/j.ymeth.2019.03.026.A

75. Meisner J, Goldberg JB. The Escherichia coli rhaSR-PrhaBAD inducible promoter system allows tightly controlled gene expression over a wide range in Pseudomonas aeruginosa. *Appl Environ Microbiol*. 2016;82(22):6715-6727. doi:10.1128/AEM.02041-16

76. Choi KH, Schweizer HP. mini-Tn7 insertion in bacteria with single attTn7 sites: Example Pseudomonas aeruginosa. *Nat Protoc*. 2006;1(1):153-161. doi:10.1038/nprot.2006.24

77. Di Salvo M, Puccio S, Peano C, Lacour S, Alifano P. RhoTermPredict: An algorithm for predicting Rho-dependent transcription terminators based on Escherichia coli, Bacillus subtilis and Salmonella enterica databases. *BMC Bioinformatics*. 2019;20(1):1-11. doi:10.1186/s12859-019-2704-x

Figures

Figure 1: 1st step motility rescue mutations in PFLU1131 A: Diagram of PFLU1131 protein, indicated positions of 1st step motility-rescuing mutations. Predicted protein domains (pfam) are indicated in lilac, amino acid positions are indicated as numbers below each domain. Mutations are indicated in blue boxes, with the number of replicate lines gaining each mutation indicated in brackets. **B)** FleQ-homolog encoding gene PFLU1132 is essential for rescued flagellar motility in the *PFLU1131*-del15 mutant strain and depends on the presence of the kinase mutation. Scale bars (white) = 12.6mm. Transcription factor gene PFLU1132 is deleted and reintroduced as a single-copy chromosomal insertion expressed from an L-rhamnose inducible promoter system (*rhaSR-PrhaBAD*). The same complementation lacking the *PFLU1131*-del15 mutation, as well as an empty expression system transposon were included as further controls. Photographs of motility after 1 day incubation in 0.25% agar LB plates supplemented with or without 0.15% L-rhamnose for induction of transcription factor expression. **C)** Volcano plots indicating impact of *PFLU1131*-del15 mutation on the transcriptome relative to AR2. Red points indicate significantly differentially expressed genes. Triangles indicate flagellar genes, Squares indicate PFLU1131/2 and adjacent genes, and Circles indicate all other genes.

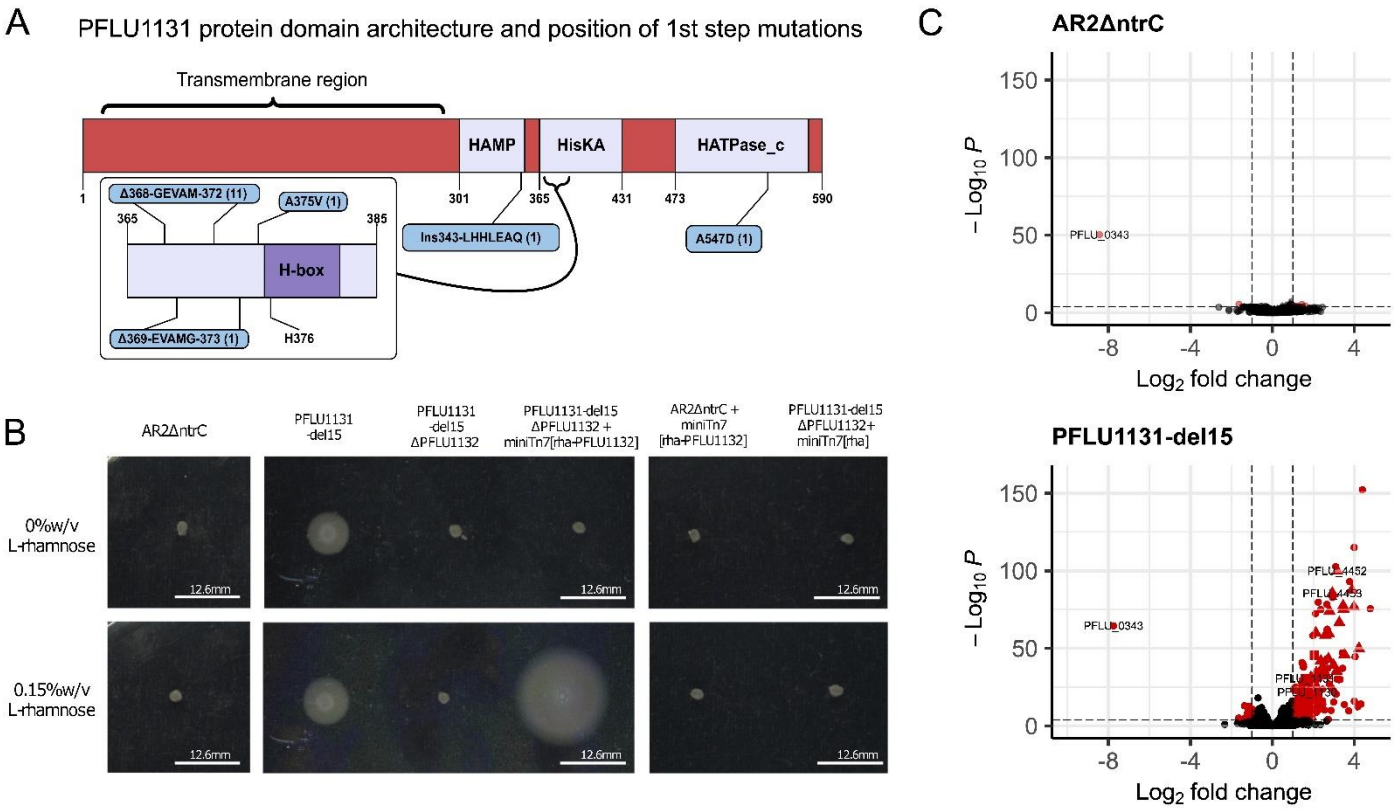


Figure 2: Log2Fold changes in gene expression for RpoN-EBP controlled genes after 1st step motility mutations, relative to AR2 ancestor. AR2 ΔntrC and AR2 ΔntrC 1st step differentially expressed genes were determined by RNA sequencing (this work). AR2 1st step differentially expressed genes were identified by RNA microarray and originally reported by Taylor *et al.* 2015. List of RpoN-EBP controlled genes was obtained from Jones *et al.,* 2007. Groups of genes by function are indicated by the coloured bars above the plot: Red – Flagellum, Blue – Nitrogen, Teal – Other/unknown.

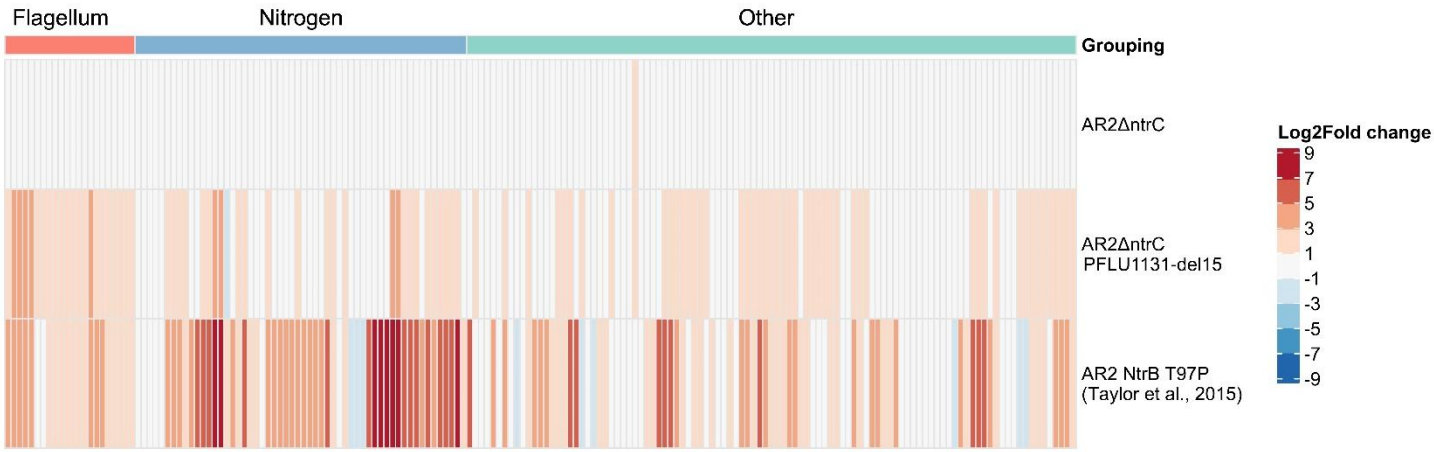


Figure 3: Phenotypes of motility rescue mutants. A) Time to emergence of motile zone (Days) for 1st and 2nd step mutants in the AR2ΔntrC and AR2 backgrounds. Percentages above each plot indicate proportion of replicate evolving that step within 6 week assay cutoff. **** indicates significant difference in mean time to emergence from all other groups (Dunn test). **B)** Race assay as measure of motility fitness. Distance moved over 24hrs in 0.25% Agar LB plates measured relative to the AR2 ntrB-T97P mutant. P-value indicated is generated from a two-sample Wilcox test. **C)** Fitness in LB broth measured as area under the 24hr growth curve (AUGC) relative to AR2 ntrB-T97P. P-values above plots generated from Dunn tests.

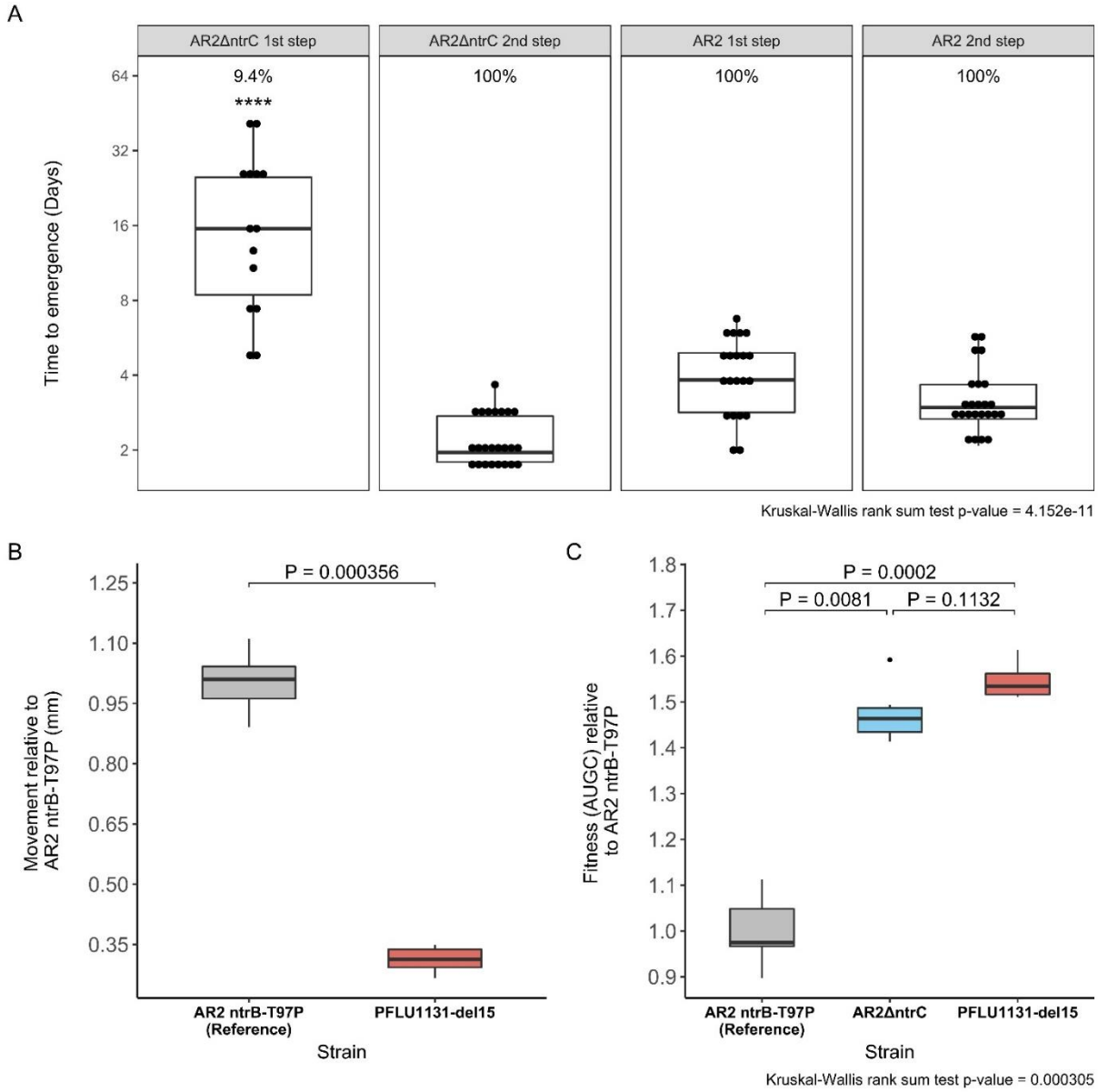


Figure 4: Diverse 2nd step mutations indicate multiple global regulatory strategies for enhancing rescue of motility through PFLU1131/2. **A)** Diagram of mutational distribution across the SBW25 chromosome for 1st and 2nd step motility mutants. Each ring represents the chromosome, with black dots indicating non-synonymous mutations. The size of each dot is proportional to the number of mutations occurring in that locus in independent replicates. Mutational targets are coloured by functional category: Red – PFLU1131/2, Purple – PFLU1583/4, Blue – Metabolic, Yellow – Global regulatory, Grey – other. **B)** Principle component analysis of transcriptomic data for ancestral, 1st step and 2nd step motility mutants. **C and D)** Volcano plots of differentially expressed genes for 2nd step mutations relative to AR2. Red points are significantly differentially expressed. Triangles indicate flagellar genes, Squares indicate PFLU1131/2 and adjacent genes, and Circles indicate all other genes. **E)** Log2Fold change in gene expression for RpoN-EBP controlled genes after PFLU1583 Δ48-74 mutation – relative to the 1st step *PFLU1131*-del15 mutant. RpoN-EBP controlled genes grouped by function are indicated by the coloured bars above the plot: Red – Flagellum, Blue – Nitrogen, Teal – Other/unknown

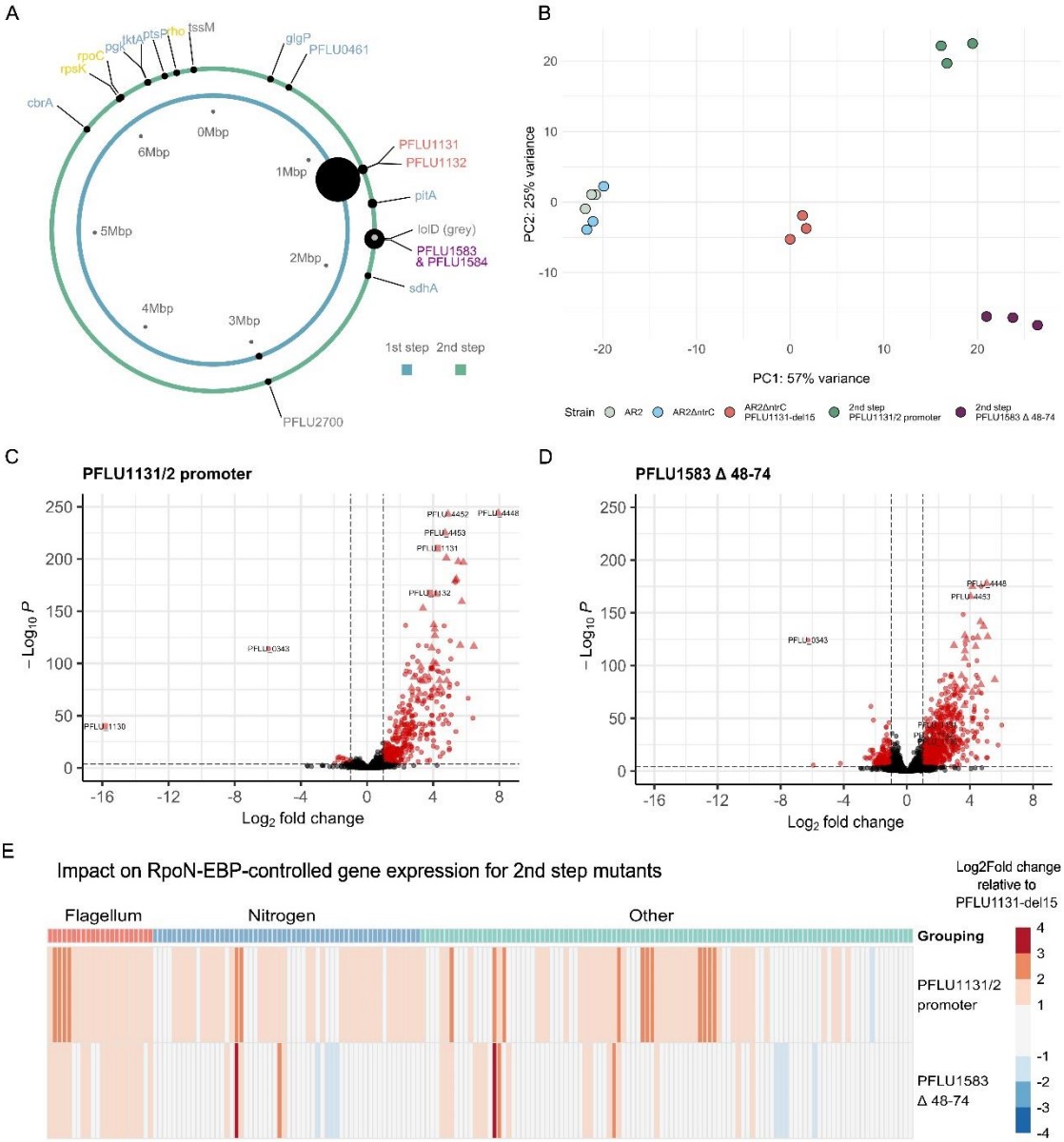


Figure 5: 2nd step PFLU1131/2 promoter mutation increases PFLU1131/2 expression and promiscuous activity. A) PFLU1130/1/2 genetic locus in the SBW25 chromosome, and impact of the 1.59kbp deletion identified in one 2nd step motile mutant. Location of the RpoN binding site upstream of PFLU1127/8/9 was predicted by Jones et al., (2007). The presence of the rho-dependent terminator was determined by use of RhoTermPredict[77] **B)** Log2Fold change in gene expression (relative to AR2) heatmap of PFLU1130/1/2 and nearby RpoN-EBP controlled genes, including PFLU1127/8/9 in the ancestral AR2ΔntrC strain, 1st and 2nd step motility mutants**.**

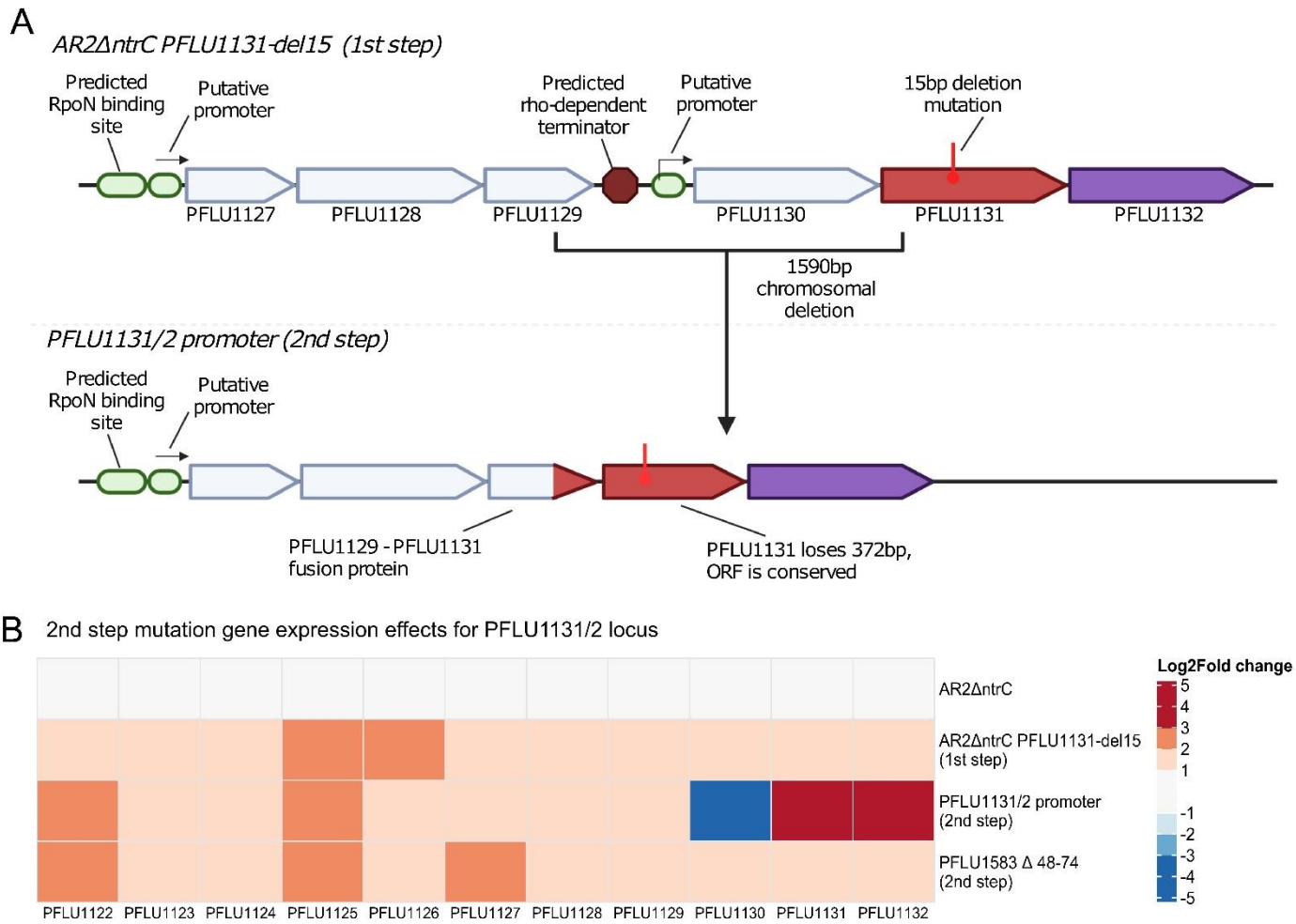
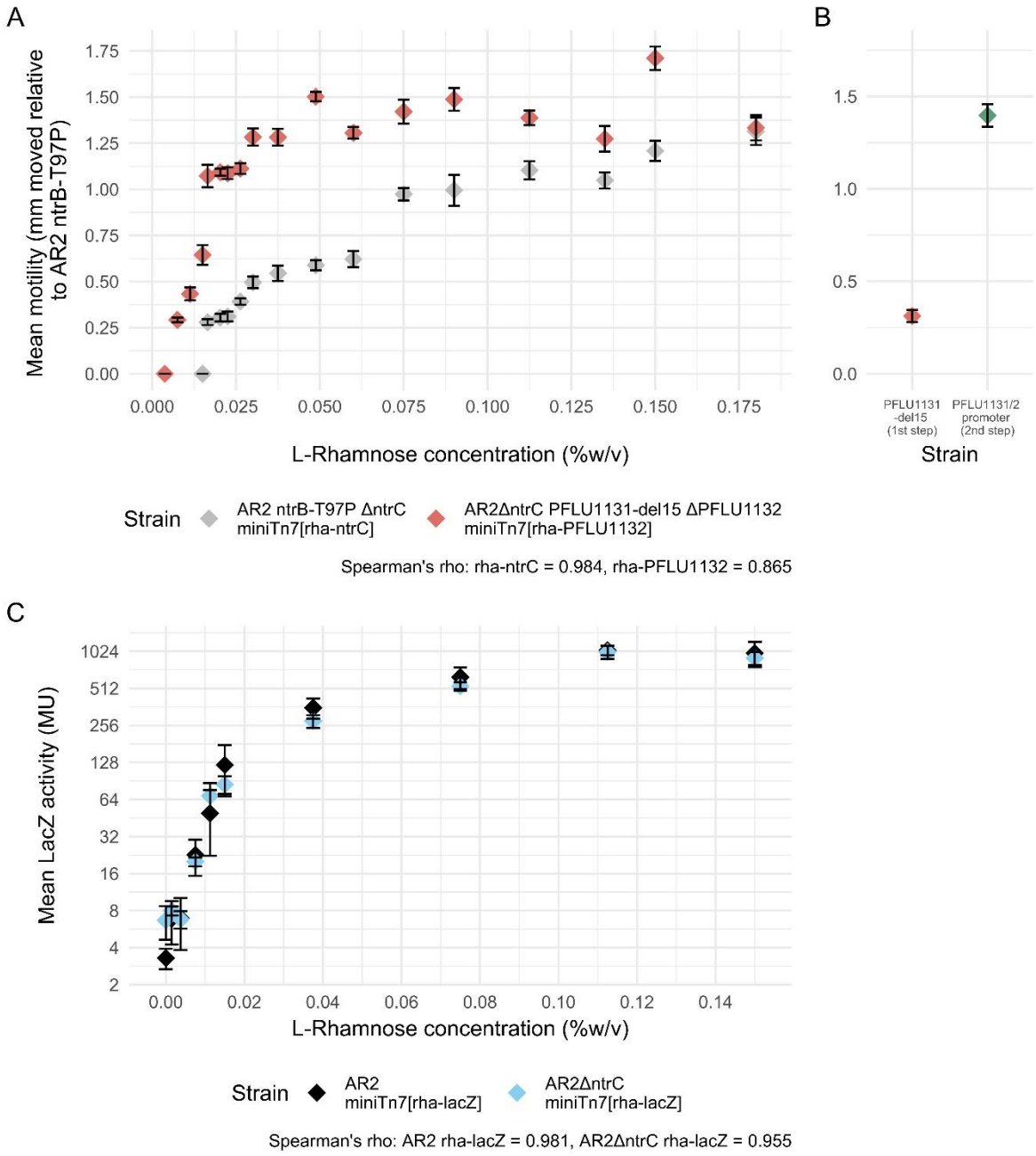
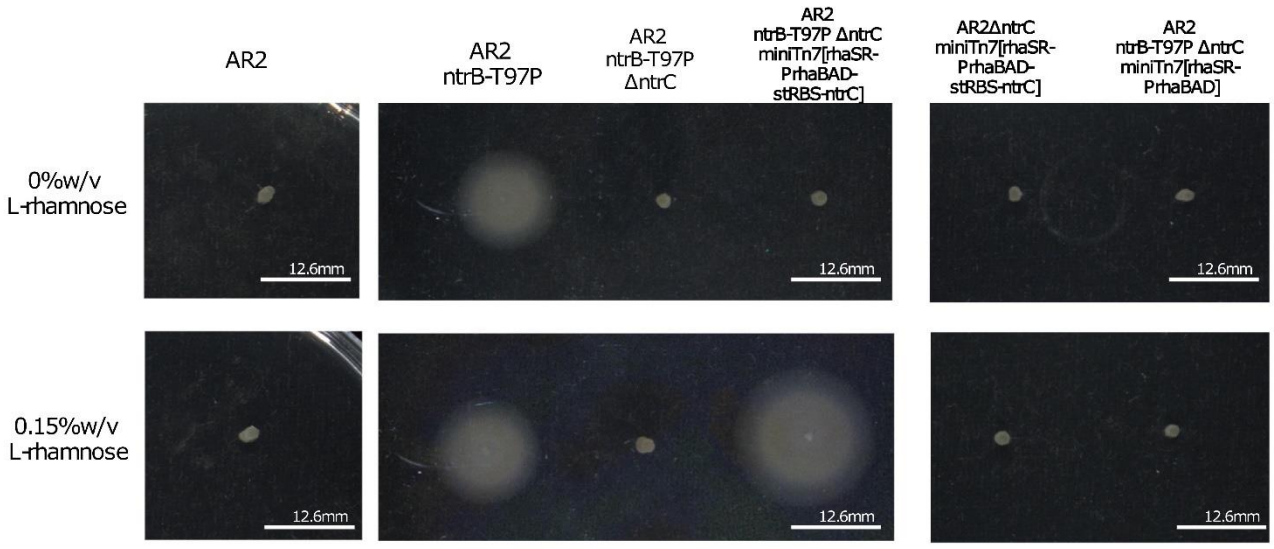


Figure 6: Impact of RpoN-EBP gene expression on motility speed. A) Motility speed (mm relative to AR2 ntrB-T97P) plotted against increasing L-Rhamnose concentration (%w/v). 1st step motility mutants AR2 ntrB T97P and AR2ΔntrC *PFLU1131*-del15 had their respective RpoN-EBPs knockout out, and reintroduced on a miniTn7 transposon with the RpoN-EBP under control of the PrhaBAD and rhaSR rhamnose inducible expression system. In full, these were – AR2 ntrB-T97P ΔntrC miniTn7[rhaSR-PrhaBAD-stRBS-ntrC] (rha-ntrC) and AR2ΔntrC *PFLU1131*-del15 ΔPFLU1132 miniTn7[rhaSR-PrhaBAD-stRBS -PFLU1132] (rha-1132). **B)** Motility speed (mm relative to AR2 ntrB-T97P) of *PFLU1131*-del15 1st step and PFLU1131/2 promoter 2nd step mutants. **C)** Activity of LacZ expression reporter (Miller units – MU) with increasing L-rhamnose concentration (%w/v) for the control genetic background AR2 and AR2Δ*ntrC* with miniTn7[rhaSR-PrhaBAD-stRBS-LacZ].

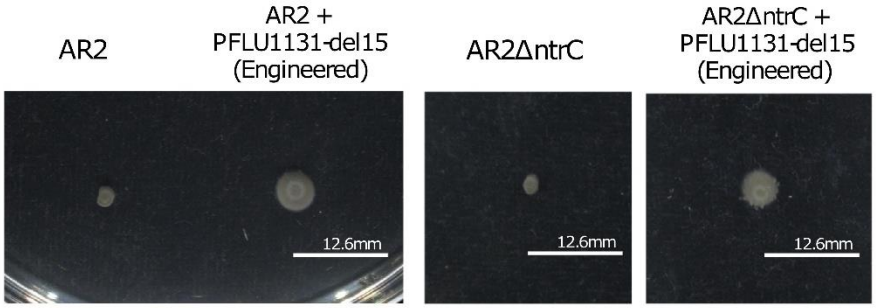


Supplementary figures

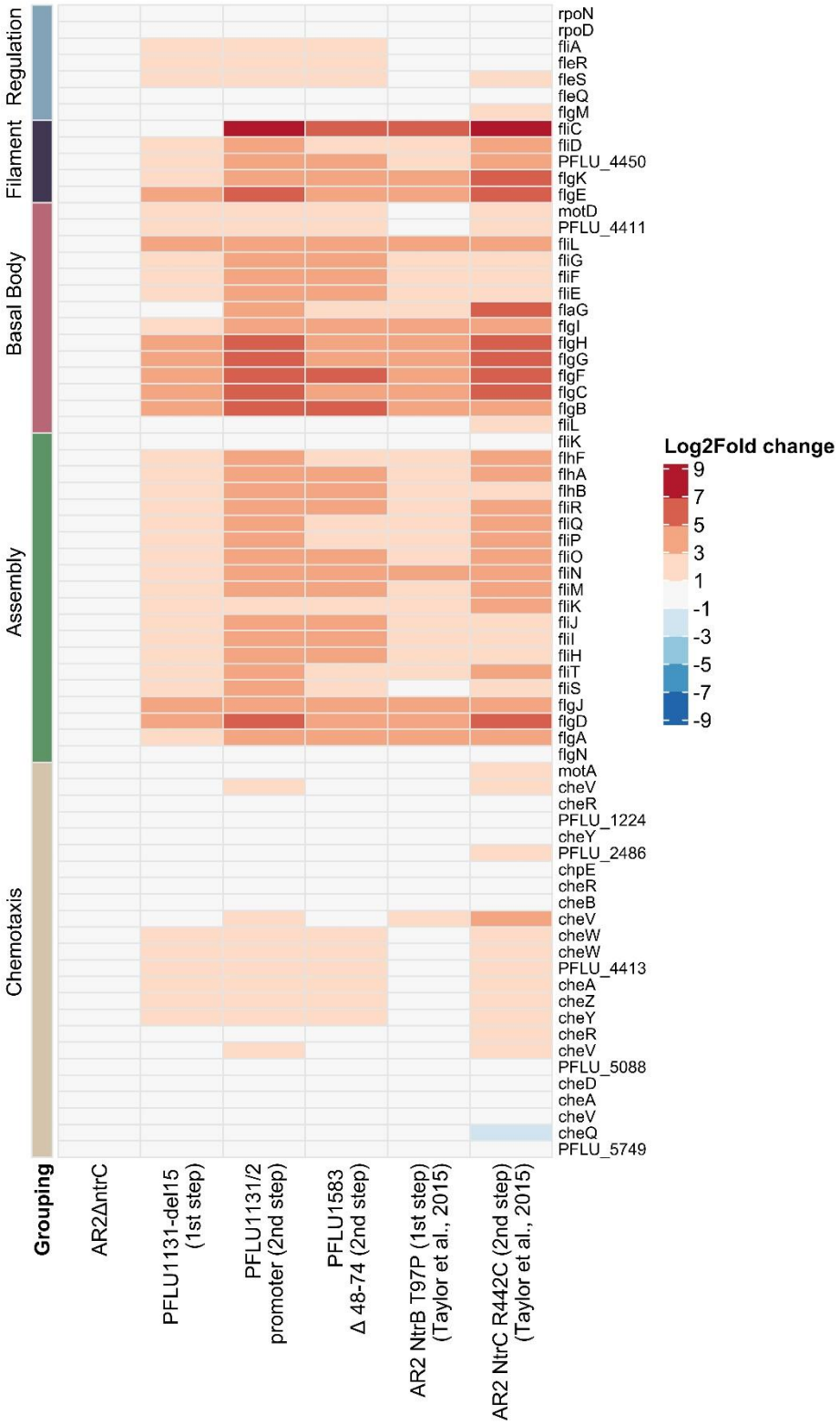
***Supplementary FigS1**:* FleQ-homolog encoding gene ntrC is essential for rescued flagellar motility in ntrB mutant strain and depends on the presence of a kinase mutation. Transcription factor gene NtrC was deleted, and then reintroduced as a single-copy chromosomal insertion expressed from an L-rhamnose inducible promoter system (*rhaSR-PrhaBAD*). The same complementation lacking the ntrB-T97P mutation, as well as an empty expression system transposon were included as further controls. Photographs of motility after 1 day incubation in 0.25% agar LB plates supplemented with or without 0.15% L-rhamnose for induction of transcription factor expression.



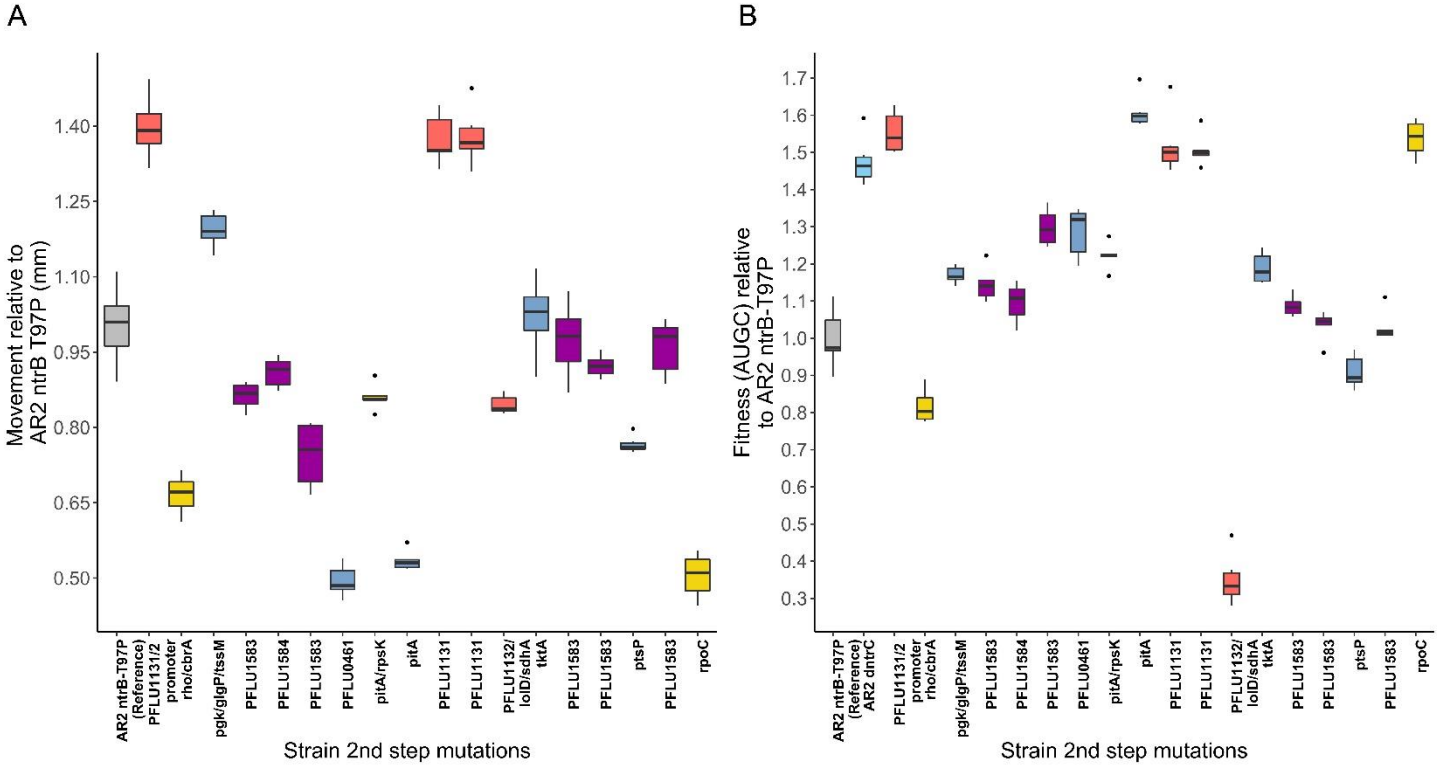
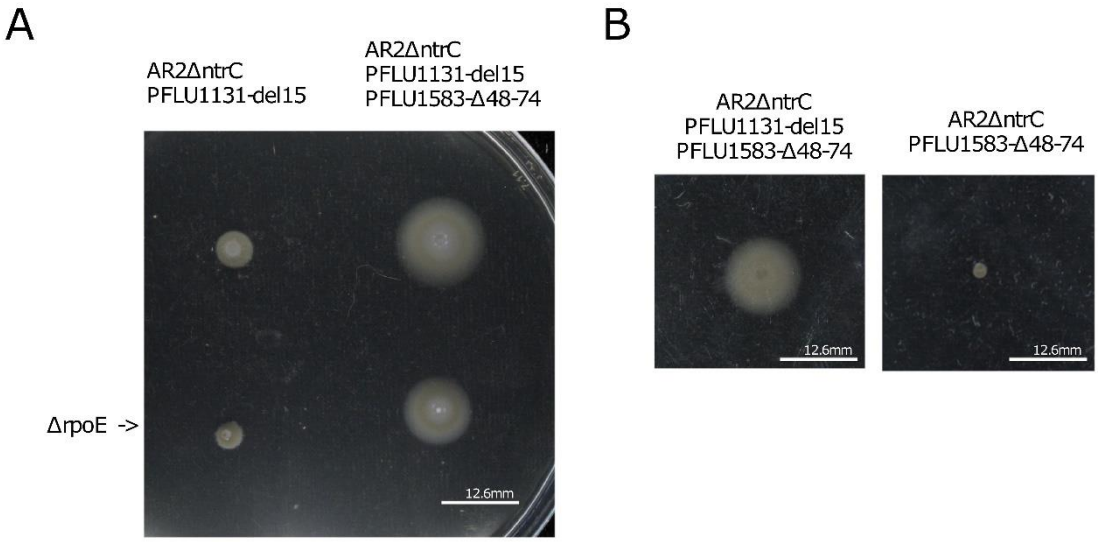
Supplementary Figure S2: PFLU1131-del15 mutation engineered into AR2 and AR2ΔntrC ancestral lines provides motility in both conditions. Indicates that rescue of motility by this mutation does not depend on Δ*ntrC* and is the sole mutation required to do so.



Supplementary Figure S3: Impact of motility rescue mutations of flagellar and chemotaxis gene expression. Log2Fold changes in gene expression relative to AR2 are shown for all genes associated with flagellar motility in *P. fluorescens* SBW25. Functional groups of genes are indicated by the coloured bars and labels to the left of the plot.



Supplementary Figure S4: Phenotypes of motility rescue 2nd step mutants. **A)** Race assay as measure of motility fitness. Distance moved over 24hrs in 0.25% Agar LB plates measured relative to the AR2 ntrB-T97P mutant. **B)** Fitness in LB broth measured as area under the 24hr growth curve (AUGC) relative to AR2 ntrB-T97P. In both plots, mutations are coloured by functional category: Red – PFLU1131/2, Purple – PFLU1583/4, Blue – Metabolic, Yellow – Global regulatory, Grey – other.



Supplementary Figure S5: PFLU1583 Δ48-74 mutation depends to presence of PFLU1131 mutation, and does not act through sigma factor RpoE. **A)** Motility of PFLU PFLU1583 Δ48-74 with and without accompanying *PFLU1131*-del15 mutation. **B)** *rpoE* knockout does not revert PFLU1583 mutant to 1st step motility phenotype.