

Evolving biocomplexity

Experimental studies on the evolution of protein complexes and ecological diversity

Flohr, Regis

DOI

[10.4233/uuid:27f84d52-31cb-47ec-bc57-c4f30b6b1c5b](https://doi.org/10.4233/uuid:27f84d52-31cb-47ec-bc57-c4f30b6b1c5b)

Publication date

2017

Document Version

Final published version

Citation (APA)

Flohr, R. (2017). *Evolving biocomplexity: Experimental studies on the evolution of protein complexes and ecological diversity*. [Dissertation (TU Delft), Delft University of Technology].
<https://doi.org/10.4233/uuid:27f84d52-31cb-47ec-bc57-c4f30b6b1c5b>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Evolving biocomplexity

Experimental studies on the evolution of protein complexes and
ecological diversity

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben;
voorzitter van het College van Promoties,
in het openbaar te verdedigen op vrijdag 23 juni 2017 om 12:30 uur

door

Régis Christian Emil FLOHR

Master of Science in de Life Science & Technology
geboren te Leiderdorp, Nederland

Dit proefschrift is goedgekeurd door de promotoren:

Prof. dr. A.M. Dogterom
Dr. H.J.E. Beaumont

Samenstelling promotiecommissie bestaat uit:

Rector Magnificus,	voorzitter
Prof. dr. A.M. Dogterom,	Technische Universiteit Delft, promotor
Dr. H.J.E. Beaumont,	Technische Universiteit Delft, copromotor

Onafhankelijke leden

Prof. dr. C. Dekker,	Technische Universiteit Delft
Prof. dr. J.A.G.M. de Visser,	Wageningen University
Prof. dr. K. Thormann,	Justus-Liebig-Universität Giesen, Germany
Dr. D.E. Rozen,	Universiteit Leiden
Dr. R.T. Dame,	Universiteit Leiden
Prof. dr. J.T. Pronk,	Technische Universiteit Delft, reservelid



Bionanoscience Department
Think big about life at the smallest scale



Keywords: compositional evolution, bacterial flagellar motor, adaptive radiation

Printed by: Proefschriftmaken.nl

Front & Back: The front cover displays evolution taking place in a petridish containing semi-solid agar (chapter 2-5; photo by Thierry Janssens). The back cover shows a biofilm produced by one of the genotypes (WS) playing an essential role in chapter 6 of this dissertation.

Copyright © 2017 by R.C.E. Flohr

Casimir PhD series, Delft-Leiden 2017-10

ISBN 978-90-8593-294-9

An electronic version of this dissertation is available at:
<http://repository.tudelft.nl/>.

Table of contents

1	INTRODUCTION	1
1.1	ORIGINS OF BIOLOGICAL COMPLEXITY.....	1
1.2	EVOLUTION OF COMPLEXITY: PROTEIN COMPLEXES	4
1.3	EVOLUTION OF COMPLEXITY: ECOLOGICAL DIVERSITY.....	11
1.4	THESIS OUTLINE.....	14
1.5	LITERATURE	16
2	EXPERIMENTAL DEMONSTRATION OF EVOLUTIONARY INTEGRATION OF ORTHOLOGOUS COMPONENTS INTO A PROTEIN COMPLEX	25
2.1	ABSTRACT	25
2.2	INTRODUCTION.....	26
2.3	RESULTS.....	28
2.4	DISCUSSION.....	51
2.5	ACKNOWLEDGEMENTS.....	60
2.6	MATERIALS AND METHODS	60
2.7	LITERATURE	66
2.8	SUPPLEMENTARY INFORMATION	71
3	THE MUTATIONS UNDERLYING COMPOSITIONAL EVOLUTION	73
3.1	ABSTRACT	73
3.2	INTRODUCTION.....	74
3.3	RESULTS.....	75
3.4	DISCUSSION.....	98
3.5	CONCLUSION	101
3.6	ACKNOWLEDGEMENTS.....	102
3.7	MATERIALS AND METHODS	102
3.8	LITERATURE	104
3.9	SUPPLEMENTARY INFORMATION	108
4	MUTATIONAL AND PHENOTYPIC TRAJECTORIES OF FOREIGN COMPONENT INTEGRATION IN THE BACTERIAL FLAGELLAR MOTOR	117
4.1	ABSTRACT	117
4.2	INTRODUCTION.....	118
4.3	RESULTS.....	119
4.4	DISCUSSION.....	134
4.5	CONCLUSION	137
4.6	MATERIALS AND METHODS	138

4.7	LITERATURE	140
5	DIVERSITY OF FIRST-STEP ADAPTIVE MUTATIONS IN FOREIGN COMPONENTS DURING COMPOSITIONAL EVOLUTION	145
5.1	ABSTRACT	145
5.2	INTRODUCTION.....	146
5.3	RESULTS.....	147
5.4	DISCUSSION.....	155
5.5	CONCLUSION	158
5.6	ACKNOWLEDGEMENTS.....	159
5.7	MATERIALS AND METHODS	159
5.8	LITERATURE	160
5.9	SUPPLEMENTARY INFORMATION	164
6	FOUNDER NICHE CONSTRAINS EVOLUTIONARY ADAPTIVE RADIATION	165
6.1	ABSTRACT	165
6.2	INTRODUCTION.....	166
6.3	RESULTS.....	167
6.4	DISCUSSION.....	173
6.5	ACKNOWLEDGEMENTS.....	174
6.6	MATERIALS AND METHODS	175
6.7	LITERATURE	177
	SUMMARY	181
	SAMENVATTING.....	185
	DANKWOORD	189
	CURRICULUM VITAE	191
	LIST OF PUBLICATIONS	193

1 Introduction

1.1 Origins of biological complexity

Life on Earth is overwhelmingly diverse and complex. Living systems display an astonishing variety of structures and function over an extremely broad range of temporal and spatial scales: from the hydrolysis of an ATP molecule in a muscle cell to the evolution of a grassland ecosystem over billions of years. One of the biggest challenges in biology is to answer the following question: how did this complexity come about?

The search for an answer to this question was helped tremendously by the fact that biological complexity can be ordered in two powerful ways. First, the functional structures of living systems can be decomposed hierarchically on the basis of the size scale at which lower-level components can be identified. This hierarchical view of life gives rise to a network of nested functional structures in which molecular events, such as the hydrolysis of an ATP molecule in a muscle cell, are connected to, for example, the dynamics of a species community in grasslands (Figure 1.1). As such, it provides a single framework within which one can consider how biological complexity functions and emerges from lower levels of organization.

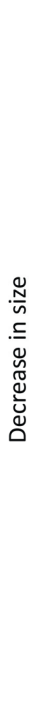
Second, the different kinds of functional structures that can be recognized at a single level of biological organization, can be classified in hierarchical groups on the basis of their structural similarities. One early example of this is the hierarchical classification of species, which was formalized for the first time by Linnaeus. The resulting nested clusters of species have a tree like structure and formed the basis for the evolutionary hypotheses put forward by Darwin and others. The evolutionary theory that sprung from this unifies all life that ever existed on Earth in a single evolutionary tree (discussed in more detail below). We now know that structural units at other levels of biological organization than that of species can also be classified in a tree-like hierarchy based on structural resemblance (e.g. proteins). This evolutionary view of life provides a single framework, within which one can consider how biological complexity arose by evolution.

Together, recognition of life's hierarchical structure and tree-like evolutionary history have allowed science to begin to make sense of the origins of biological complexity. However, most of what is known about the mechanisms that underpin the evolution of complexity and diversity comes from comparative studies. Direct experimental insight into the transformations that are indicated by this work is still very limited. The work described in this thesis sets out to contribute to the relatively young but growing body of work that tests and expands our mechanistic understanding the origin of biological complexity using experiments. It focuses on the evolution of functional structures at two disparate levels of biological organization: protein complexes at the molecular level and ecological diversity at the community level. In both cases, bacterial populations were used to experimentally explore a key mechanism behind the evolution of complexity.

1.1.1 Heritability and natural selection

The basis of current evolutionary thought, with the main concepts of heritability and natural selection, was laid down at the end of the 18th century. Erasmus Darwin (1731-1802) published in 1794 his early ideas on the theory of evolution. In his theory, he introduced the concept of heritability [1], which holds the idea that characteristics of the parent are passed on to offspring. Around that same time, James Hutton (1726-1797) introduced the concept of natural selection, by which characteristics of a population change over time due to the interaction between individuals of that population and the environment [2]. As natural selection would favour individuals with an increased ability to survive and reproduce in a certain environment (i.e. increased fitness), the presence of characteristics of these individuals would increase in frequency in the population. Jean-Baptiste Lamarck (1744-1829) continued developing these ideas and hypothesized that acquired traits of the parent, which it acquired during its lifetime by using it in a certain environment, were passed on to offspring and thus led to gradual change over several generations. However, it was Charles Darwin, grandson of Erasmus Darwin, and his famous work *On the Origin of Species* (1859) [3] that was most influential and more correct. Although Alfred Wallace (1823-1913) independently came to similar conclusions as Darwin, Wallace's work is less well known. Darwin hypothesized that populations contain heritable variation in traits and that this initial variation (instead of variation in traits that was acquired during an organisms lifetime, such as Lamarck proposed) could be selected for by natural selection, which depends on the environment in which the population lives. In other words, a certain trait of a population gradually changes over time, because fitter organisms (fitter due to a certain variation of that trait) leave more offspring than less fit organisms. In the absence of the generation of more variation, the variation in a certain trait is thus purged until only the trait that leaves organisms most fit is retained.

With Charles Darwin's work, a plausible mechanism behind evolution was proposed, but a mechanism underlying heritability was still lacking. It was Gregor Mendel (1822-1884), nowadays known as the father of genetics, who in retrospect found this mechanism by showing the presence of certain factors that could predict phenotypical traits in pea plants. Although published in 1866 [4], his work was re-discovered at the start of the 20th century. In the 1930s it was combined with the theory of natural selection, which ultimately resulted in the Modern Synthesis, a term coined by Julian Huxley (1887-1975) in his book *Evolution: The modern synthesis* (1942) [5]. At this point, heritable units were termed genes [6] and were known to encode proteins [7], but the identity of the heritable material was still a mystery. This identity remained hidden until 1952, when Hershey and Chase confirmed that DNA was the biological entity responsible for heritability [8]. Since DNA was discovered in 1869 by Friedrich Miescher [9], much was known already about its structure, including that it consisted of four bases [10] and that not all bases were represented equally [11].



Biosphere	→	Group of ecosystems	e.g. Earth
Ecosystem	→	Group of communities and non-living environment	e.g. Pride of lions living together with herd of wildebeast in area with rivers and grassland
Community	→	Group of populations	e.g. Pride of lions living together with herd of wildebeast
Population	→	Group of single-cellular or multi-cellular organisms	e.g. Pride of lions, herd of wildebeast, population of bacteria
Multi-cellular organism	→	Group of tissues, organs, etc.	e.g. Lion, wildebeest
Tissue, organ, etc.	→	Group of cells	e.g. Skin, muscle
Cell/Single-cellular organism	→	Group of molecules	e.g. Skin cell, muscle cell / bacterium
Molecule	→	Group of atoms	e.g. DNA, RNA, protein
Atom			e.g. Carbon, nitrogen, oxygen

Figure 1.1. Biological organisation from macroscale to microscale. With each level, from biosphere to atoms, the size of structures decreases.

1.1.2 A simplified example of evolution

With above knowledge in mind, the simplest view of evolution of a population (e.g. wildebeast) can be summarized as follows. A mutation in the DNA leads to a change in a specific protein (e.g. muscle protein). This change in protein on its turn, percolating through levels of biological organization (Figure 1.1), leads to an increase in fitness of the organism harbouring the mutation. Fitness here is defined as the ability to survive and reproduce, and depends on the environment. In our example, the change in muscle protein leads to stronger muscles, which makes the bearer faster and able to escape from lions. As this will lead to a lower chance of getting killed and thus more time to reproduce, fitness increases. Because competition between individuals with and without the mutation (both ancestor and offspring) will lead to a relative increase in the number of individuals that harbour the mutation (wildebeast with the mutation tend to live longer and generate more offspring), the population will evolve from 100% non-mutated to 100% mutated. At this point, the population is said to have evolved, because the entire population has become faster.

However, evolution often is not that straightforward. In many cases, there will be interactions between organisms of several species and not just two. Furthermore, the environment tends to change, which might influence different species differently, and accidents leading to the loss of favourable genes (an example of genetic drift [12]) might occur. If we zoom in to the sub-organism level, we find more factors complicating evolution. Epistasis, defined as the phenomenon that

phenotypic consequences of a mutation are dependent on the genetic background [13,14], might shape evolution. It goes beyond the scope of this thesis to discuss every aspect of evolution in detail, but it goes without saying that evolution is not that simple as in the example illustrated above.

In this thesis we focussed on the evolution of functional structures at two disparate levels of biological organization: protein complexes at the molecular level and ecological diversity at the community level. For protein complexes, it is not exactly clear how they could have evolved. The issue revolves around the need for selectively-accessible, step-wise mutational trajectories, the existence of which appears to be, at first sight, unlikely for such complex systems. For ecological diversity, the issue is that theory suggests that the properties of the ancestral species of a range of related, ecologically diverse organisms (e.g. as is the case for Darwin's finches) has an impact on the process of evolutionary diversification. In order to demonstrate this connection experimentally, one needs to disentangle the role of the ecology of an organism from the role of its capacity to generate heritable phenotypic variation (*i.e.* its evolvability).

1.2 Evolution of complexity: protein complexes

Protein complexes are composed of multiple, different protein components that together perform a specific function. Examples include the NADH:ubiquinone oxidoreductase [15], the V-type ATPase [16,17] or the bacterial flagellar motor (BFM) [18]. Comparative genomics studies suggest that protein complexes gradually evolved by the addition of pre-existing protein parts to simpler assemblies [15,19–25]. Two arguments render the existence of protein complexes difficult to explain by evolution. The first argument revolves around the apparent presence of many strong and specific interactions (epistasis) between the protein components in these complexes [26–30]. When there are many strong and specific interactions between components of a protein complex, improving function by point mutations, and thus gradually and step-by-step, in one component would need many mutations in different components simultaneously. As the number of mutations increases, this becomes less probable. In this view, the existence of step-wise, selectively-accessible mutational trajectories that can forge new functional interactions between pre-existing proteins is highly unlikely. The second argument is that protein complexes contain components that are essential to their function [19]. This fact is used by some to argue that protein complexes are “irreducibly complex” and therefore cannot have evolved [19,31]. However, counter arguments to the arguments presented above can be provided. The first argument can be countered by raising the possibility of *modularity*, which is the presence of a low number of epistatic interactions between components (in this case called *modules*) in a protein complex [26,29]. If this were to be true, the functionality of modules would be relatively independent of other modules and components would have a degree of structural and functional independence. Furthermore, modules would have little influence on the functioning of modules outside of themselves [29]. In other words, when protein complexes have a degree of functional modularity, they interact towards the higher-order function of the complex in a less strong and specific way than might appear to be the case at first sight. As such, modularity would open up adaptive trajectories of mutational steps that are individually adaptive, allowing fine-tuning of a protein complex. In addition, modularity would allow the evolutionary origin of protein complexes by step-wise integration of pre-existing proteins. This mode of evolution, which is termed *compositional evolution*, is explained in more detail in the next section. The argument of “irreducible

complexity” can be countered by the principle of *evolutionary lock-in*, which means that initially non-essential components can become essential by neutral or adaptive evolution [17,28]. As these counter arguments make the evolution of protein complexes more likely, this brings forward the question of how the protein complexes that we witness today, came to be in the first place.

1.2.1 Compositional evolution

The most conventional thought on evolution in general is Darwinian evolution: evolution by natural selection and small steps, due to point mutations creating slight variations of a certain characteristic. Although this is a plausible mechanism for the evolutionary origin and adaptation of single proteins (or multimers thereof), such as for example enzymes [32], the evolution of large protein complexes by only such gradual evolution is less straightforward to explain, as discussed above. Additionally, if we consider that they often contain components of which homologues are found to perform other functions in the cell (some examples in the BFM are mentioned in [19]), this hints at additional mechanism of evolution.

Experimental insight into the mechanisms by which multi-component protein complexes evolved is scarce, but comparative genomics studies suggest that protein complexes evolved by the step-wise addition of pre-existing protein components to more simple assemblies [15,19–25], a process called *compositional evolution* [33]. A specific example for the bacterial flagellar motor (BFM) is found in *Shewanella oneidensis* MR-1, for which comparative genetics analyses have indicated that one of two stator complexes was acquired by HGT [34] (for more details on the BFM, see below). In compositional evolution, the source of *foreign components*, which includes foreign genes and the proteins they encode, can be either extracellular or intracellular and mechanisms that provide them include horizontal gene transfer (HGT) [30,35–42] or gene duplication [43–45], respectively. Furthermore, for compositional evolution to be possible, it requires either immediate compatibility of the foreign components (in terms of expression of genes and functionality of proteins) or the capacity of evolution to cause compatibility.

Recently, an experimental study provided direct experimental evidence of compositional evolution. In an elegant study [17], they showed, by using ancestral gene resurrection and mutation reconstruction, that complexity of the ring of V-type ATPases in yeast could have increased due to gene duplication and divergence of a specific ring component. Other experimental studies that have examined mechanisms behind compositional evolution, include studies that investigated the consequences of replacing genes that encoded protein-complex components for homologous genes [39,42] or studies that examined protein oligomerization (e.g. [46,47]). Although above studies have provided experimental insight into compositional evolution, additional experimental data is necessary to further establish this fundamental process in the evolution of protein complexes. In chapters 2-5 of this thesis, we will examine the creative potential of and mechanisms of compositional evolution in more detail, by following the evolutionary incorporation of pre-adapted, orthologous components into a protein complex.

1.2.2 Compositional evolution and the fitness landscape

Protein complexes are characterised by the presence of epistatic interactions between protein components. Because of these interactions, the effect on organismal fitness of a mutation in one component might be influenced by the state of another component. Because such interactions

suggest that similar mutations can have different fitness effects that depend on the genetic background, epistasis might influence evolution and shape the *fitness landscape* [14,48–50].

A fitness landscape is a metaphor that illustrates the effects of mutations on organismal fitness (Figure 1.2) and is represented by all possible genotypes in the xy-plane (genotype space) and the organismal fitness corresponding to the genotypes on the z-axis. When there is only one fitness optimum possible, natural selection will drive the population up the single peak by point mutations along an evolutionary path (Figure 1.2A). However, when epistasis occurs and especially sign epistasis, which means that a mutation will have either a positive or negative effect on fitness, depending on the genetic background [49], the fitness landscape will show multiple peaks separated by valleys (Figure 1.2B; rugged fitness landscape). If natural selection drives the population up the low peak (blue path), this population will never reach the global optimum by point mutations, because the path towards the global optimum leads through a fitness valley (yellow path). As a decrease in fitness is disfavoured by natural selection, the population is said to be evolutionarily trapped. However, when natural selection would have followed another evolutionary path from the start, which consisted of many steps of small increase in fitness (green path), the global optimum would have been available by just point mutations.

Compositional evolution by definition represents a jump in genotype space and is analogous to many mutations at once. When a foreign component is added, the jump will result in a position in genotype space outside of the current genotype space, as the genotype is enlarged, and thus allows for new genotype territory to be explored. This territory would have been unavailable if point mutations were the only type of mutation possible. When a gene that encodes a component of a protein complex is swapped for a homologous gene, for instance acquired through HGT, the swap could lead to a jump in the current genotype space. In case of a rugged fitness landscape (Figure 1.2B), compositional evolution could thus lead to a jump from the low fitness peak to the high fitness peak, thereby liberating a population that was evolutionarily trapped and opening up new possibilities for future adaptation.

In summary, compositional evolution can facilitate evolution by providing access to possibilities not reachable via natural selection through point mutations. However, if epistatic interactions are strong and specific, compositional evolution might not be possible due to the need for many mutations in other components to allow for interactions with the foreign component.

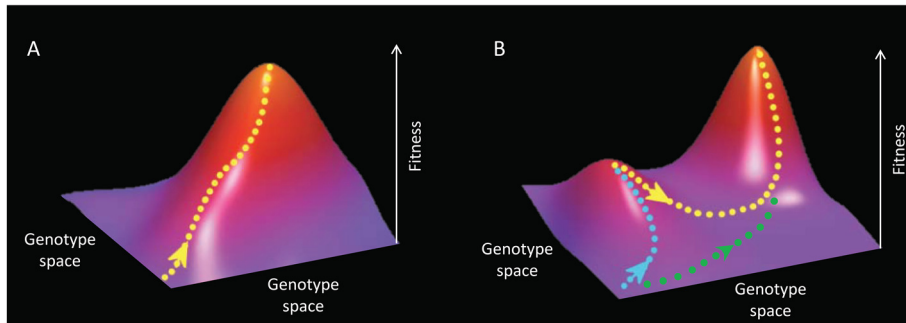


Figure 1.2 Diversity in fitness landscapes. Evolutionary paths are indicated by coloured dots. **A.** Single-peaked fitness landscape. All point mutations lead to an increase in fitness. **B.** Multi-peaked fitness landscape. Although point mutations lead to an increase in fitness (blue path), this population cannot reach the highest peak (yellow path), as this would require mutations that decrease fitness. Mutations along the green path would allow a population to reach the highest peak. Adapted from Poelwijk et al. 2007 [14].

1.2.3 Probing the dynamics of compositional evolution at the compatibility horizon

Before introducing the model system that we used in this study, let us consider a hypothetical situation in which compositional evolution can lead to the co-option of any foreign component in the biosphere. If this were to be true, the creative potential of compositional evolution would be constrained by three factors. The first factor that constrains the creative potential of compositional evolution is the functional diversity of foreign components that are present in the biosphere. Depending on the initial function of the foreign component, compositional evolution leads to specific phenotypic effects. For instance, co-option of a protein with an ATPase function will lead to different effects than co-option of a protein with DNA-binding properties. Thus, the more diversity in initial function of foreign components, the more creative can compositional evolution be. The second factor that constrains the creative potential of compositional evolution is the subset of foreign components that is immediately capable of interacting functionally with the host. *Functional interaction* is defined here as the expression of genes and interaction of the corresponding proteins with each other and other cellular systems. Furthermore, functional interaction means that co-option of the foreign component results in phenotypic effects that are selectable. Thus, the more foreign components that show functional integration immediately, the more creative can compositional evolution be. The third factor that constrains the creative potential of compositional evolution is the capacity of evolution to forge functional interactions between the remaining foreign components in the biosphere and the host. Foreign components that did not have a functional interaction immediately, but can be functionally integrated can expand the range of foreign components that is amenable for compositional evolution. As such, it can increase the creative potential of compositional evolution. Together, these three factors divide the foreign components that can be used for compositional evolution into three domains. The first domain is composed of foreign components that are compatible immediately and can be modified and functionally integrated by evolution. The second domain is composed of foreign components that are incompatible initially, but can be modified and

functionally integrated by evolution. Finally, the third domain is composed of foreign components that are incompatible initially and cannot be modified and functionally integrated by evolution within a defined interval of evolutionary time. In other words, the creative potential of compositional evolution is bound by what may be considered a *compatibility horizon* from which beyond no foreign components can be co-opted for compositional evolution.

In this thesis, we begin to probe the dynamics of compositional evolution at the compatibility horizon. One issue that we address is the incorporation of non- or sub-optimal, orthologous components. Is evolution capable of integrating or improving the integration of these components? If so, where in the genome are the mutations located? And what would the mutational trajectories to (improved) incorporation look like? To answer these and more questions, we need a model system that allows us to monitor the phenotypic effects of foreign components.

1.2.3.1 Model system: bacterial flagellar motor

As mentioned before, experimental insight into the mechanisms behind the evolution of protein complexes is scarce (e.g. [17,39,42,46,47]). What is needed is additional experimental insight into the creative potential of and mechanisms behind compositional evolution to further establish this fundamental process in the evolution of biological complexity.

In order to provide additional insight into compositional evolution and dynamics at the compatibility horizon, we examined systematically the consequences of replacing genes that encode components of a protein complex, with foreign genes that encode homologous protein components from other organisms. As such, we investigated the evolutionary possibilities of compositional evolution at the compatibility horizon. Amongst others, we examined the occurrence of immediate compatibility of the foreign components and the existence and identity of step-wise mutational trajectories that facilitated compatibility. The model system that we used to accomplish this is the Bacterial Flagellar Motor (BFM) of *Escherichia coli* (*E. coli*) MG1655.

The BFMs of *E. coli* and *Salmonella typhimurium* are amongst the most studied of all rotary protein-complexes, e.g. [18,51,52]. The BFM is highly conserved amongst bacteria [53] and is a beautiful example of a biological nanomachine that rotates at frequencies ranging from tens of Hz to hundreds of Hz [18]. Situated in the cell wall of both gram-positive and gram-negative bacteria, it couples ion — either H⁺ (e.g. [54–56]) or Na⁺ (e.g. [57]) — translocation to flagellar rotation and torque generation, which allows cells to move through their environment.

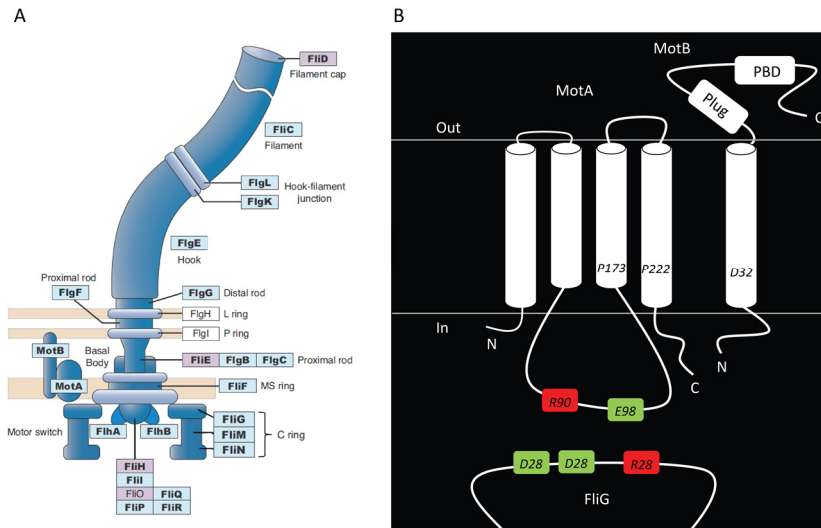


Figure 1.3. The Bacterial Flagellar Motor. A. Overview of the BFM with all of its components. Adapted from [20]. **B.** Close-up of the stator-rotor interaction site. Transmembrane domains are represented by cylinders. Horizontal lines indicate the cell membrane that separates the cytoplasm (in) from the periplasmic space (out). Charged residues (red: positive, green: negative) of MotA (R90 and E98) interact with oppositely-charged residues of FliG (D289, D288 and R281). Residues P173 and P222 are involved in conformational changes of MotA [58]. Residue D32 is essential for ion binding [56]. The plug domain is involved in preventing premature ion flow across the membrane and might be involved in correct positioning of the peptidoglycan binding domain (PBD) [59]. The PBD anchors the stator to the peptidoglycan.

The *E. coli* BFM (Figure 1.3A) is an H^+ -driven motor and is composed of more than 20 different protein components [52]. In its essence, it is composed of four major structural features, besides components involved in the export of flagellar proteins: the basal body, the stator complex, the flagellar filament and the hook that connects the basal body and flagellar filament [60].

The flagellar filament is a long (10-15 μ m) and thin (20nm) hollow tube [60] that protrudes from the cell and is composed of \sim 20,000 FliC monomers [61]. Counter clockwise (CCW) rotation of single BFMs results in a left-handed, supercoiled flagellar bundle behind the cell and rotation of this bundle results in propulsion through the environment [62,63]. This is called a run. Occasional clockwise (CW) rotation of single BFMs causes disruption of the flagellar bundle and results in cellular re-orientation. This is called tumbling and will be discussed further in section 1.2.3.2.

The hook that connects the filament with the rotary parts of the BFM, is composed of \sim 120 copies of the FliE monomer and its main feature is its flexibility [64]. As *E. coli* is peritrichously flagellated, i.e. flagella appear all around the cell body, flexible hooks allow flagella to form a proper bundle independent on the position of flagella.

The basal body is composed of a set of rings that spans the membrane and forms the rotary part of the BFM [65]. The rings are in order from outer- to inner- membrane the L-, P-, MS- and C-ring. Only the C-ring, with \sim 40nm the widest of all rings [53], is not embedded in one of the membranes. This C-ring, also called the switch complex, is composed of three different proteins (FliG, FliM and FliN) and plays a crucial role in torque generation [66,67] and switching the direction of flagellar rotation, thereby causing tumbling [68,69]. For torque generation, FliG interacts with MotA, which on its turn is part of the stator complex together with MotB (see below) [70,71]. For switching, FliM

and FliN interact with phosphorylated CheY [69,72], a protein which is part of the chemotaxis signalling pathway [73]. As the C-ring is attached to the MS-ring [66] and as the MS-ring is attached to the hook through the central rod [18], rotation of the basal body can be coupled to the outside flagellar filament.

The fourth major constituent of the BFM, and the primary focus of chapters 2-5 of this thesis, is the stator complex (Figure 1.3A and B). As mentioned above, the stator complex is composed of two different proteins, MotA and MotB [70,71]. MotA is 295 amino acids long, it contains four transmembrane (TM) helices and is especially important for the generation of torque through interaction with FliG [52,74–76]. MotB is 308 amino acids long. It contains the proton binding-site necessary for torque generation [56], a plug domain that prevents the flow of protons over the membrane when not bound to the BFM [59] and it anchors the BFM to the cell wall through the peptidoglycan binding domain (PBD) [77,78]. Per BFM ~11-12 stator complexes contribute individually to torque generation [79] and can move in and out from the BFM [80]. One stator complex consists of 4 copies of MotA and 2 copies of MotB, with TM helix 3 and 4 of MotA and the TM helix of MotB forming the ion-conducting channel [81,82].

The most likely model that couples proton translocation to flagellar rotation and torque generation is as follows [83]. When the stator complex binds to the BFM, the plug domains dimerize and this opens the ion-conducting channel. Protons flow through the channel and bind to residue D32 of *E. coli* MotB. The binding causes a conformational change in the stator complex, which results in MotA having a different interaction with FliG and a first power stroke. Deprotonation of residue D32 returns the stator-complex conformation to its pre-binding state and causes a second power stroke. Continuous flow of protons results in continuous power strokes and thus in continuous rotation of the BFM. This causes the flagella to rotate CCW and bundle together, which allows a cell to perform a run and to swim in a random direction.

1.2.3.2 Model system: nutrient-rich, semi-solid agar

Over the years, many experiments have been conducted in semi-solid agar, for instance to study bacterial chemotaxis (e.g. [84–86]) or to study the BFM (e.g. [87–89]). When bacterial cells are inoculated in semi-solid agar containing several nutrients, they will start consuming these nutrients in a preferred order [85]. Consumption of nutrients causes the creation of a nutrient gradient, which motile cells are able to follow, while they simultaneously consume more nutrients. This mechanism of sensing and responding to chemical stimuli from the environment is called chemotaxis [84]. It provides motile cells with a fitness advantage over non-motile cells, as they are able to move towards favourable conditions and away from less favourable conditions [90]. In the previous section we already mentioned the coupling between the BFM and the chemotaxis signalling pathway, and below we will illustrate this in more detail.

Figure 1.4 shows the chemotaxis signalling pathway (reviewed in [73]). Chemical attractants have the ability to bind to the chemotaxis receptors (methyl-accepting chemotaxis proteins; MCPs), of which five different types exist, each having affinity for one substrate [91]. In *E. coli* we find receptors for aspartate (Tar receptor), serine (Tsr), dipeptides (Tap), galactose (Trg) and oxygen (Aer). CheW links the chemotaxis receptors to CheA [92,93], which can phosphorylate itself (resulting in CheA-P) and two different proteins, CheY and CheB [94]. While phosphorylated CheY (CheY-P) is able to bind

to the switch complex of the BFM [69,95], phosphorylated CheB (CheB-P) controls the adaptation of the chemotaxis receptors by removing a methyl-group that was placed there by CheR [94,96,97]. CheZ dephosphorylates CheY-P and plays an important role in signal transduction termination [98]. So, how does this signalling pathway makes the cell respond to changes in nutrient concentration, which allows cells to move towards favourable conditions?

Let us assume we have a cell that rotates its flagella CCW and thus performs a run. When the concentration of attractant decreases, less attractant will bind to the receptors, which leads to an increased rate of CheA autophosphorylation [94]. This will increase the levels of CheY-P and CheB-P [94], which has two effects. On the one hand, CheY-P will bind to the switch complex of the BFM, which will cause the BFM to change rotational direction from CCW to CW [69,95]. The result of this change in rotational direction is a tumbling event during a run and leads to re-orientation of the cell. On the other hand, CheB-P will de-methylate the receptor, which reduces the ability of the receptor to induce CheA autophosphorylation. Together with the dephosphorylation of CheY-P by CheZ, this will lead to pre-stimulus levels of CCW rotational direction. The systems is now primed for additional changes in attractant concentration. When the concentration of attractant increases, this will effectively lead to lower levels of CheY-P and a lower tumbling frequency. Thus, the movement in the direction of increasing nutrient concentration is prolonged (prolonged runs). On average runs last for ~1 s and tumbling events ~0.1 s, with intervals of ~1 s between tumbling events [99,100].

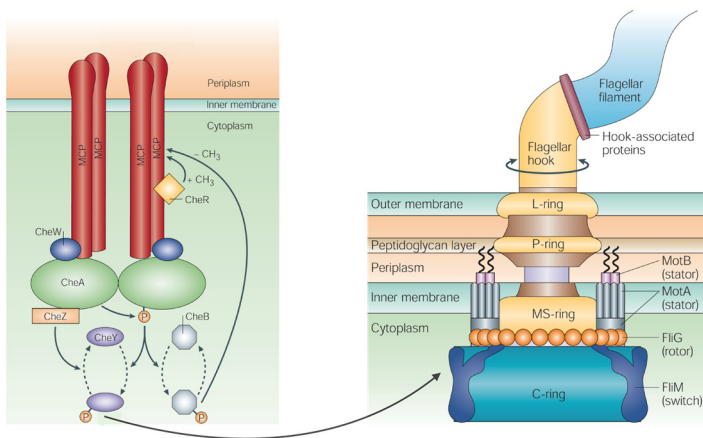


Figure 1.4. The interaction between the chemotaxis signalling pathway and the BFM. A decrease in attractant concentration results in trans-autophosphorylation of CheA. One of the resulting actions is the phosphorylation of CheY. CheY is a response regulator and binds to the switch complex of the BFM, which causes the BFM to rotate clockwise (CW) instead of counter clockwise (CCW). CW rotation causes the flagellar bundle to disrupt, which results in a tumbling event that re-orientes the cell. MCP: methyl-accepting chemotaxis proteins. Adapted from [73].

1.3 Evolution of complexity: ecological diversity

As mentioned in the introduction, the evolution of complexity at two different levels of biological organization was studied in this thesis. Besides studying the evolution of complexity at the molecular level, we also studied the evolution of complexity at the community level. In Chapter 6 of this thesis, we examined the factors that shape the evolution of ecologically diverse organisms from a single

ancestral type. In contrast to our investigations on the evolution of protein complexes, we did not focus on the underlying mutations and molecular changes. Instead, we focussed on the dynamics of diversification and their link to the properties of the ancestor.

1.3.1 Adaptive radiation

Adaptive radiation (AR), which is the rapid diversification of a single lineage into organisms with different niches, underpins the evolution of biodiversity [101–106]. Arguably the most famous example of AR, is the diversification of Galapagos finches (Figure 1.5). It was first described by Charles Darwin during his voyage on the HMS Beagle and let him to develop his ideas leading up to *On the Origin of Species* [3]. In this particular AR, it is believed that the ancestral, seed-eating ground finch, arose in South America, flew 1,000 kilometres over water and landed on the Galapagos Islands [107]. There, it found itself in the presences of many untapped resources to which it subsequently adapted. This culminated in the formation of many different species of finch with different niches (Figure 1.5).

Although there are several definitions of the concept ‘niche’, in this thesis we will use the definition following McNerny and Etienne [108–110]. Here, niche is defined as the complex of reciprocal ecological interactions between an organism and its environment, that governs organismal fitness. In the Galapagos finch example, different species feed on different resources and therefore have different interactions with the environment. Different finches thus have different niches.

Over the years, much research has been performed regarding factors that influence AR, both in extant (e.g. [103,105,111–115]) as well in experimental ARs (e.g. [116–126]). This showed that there are two prerequisites that are essential for AR to unfold: evolvability of organisms and ecological opportunity. Evolvability is the capacity of a group of organisms to generate heritable and selectable phenotypic variation [127,128] upon which natural selection can act. Without variation, offspring will not differ from its ancestor and will thus essentially be the same. Ecological opportunity on the other hand, emerges when the environment allows organisms with different niches to invade and co-exist with organisms that are already present. Together, evolvability and ecological opportunity will allow for an ancestor to diversify into a range of organisms with different niches. If we turn to the Galapagos finch example, ecological opportunity could be observed in the presence of many resources that were not used by the seed-eating ancestor, e.g. insects and fruits. Evolvability of this ancestor is apparent by the number of different finches that evolved from it and used these different resources. Without evolvability, there would have been still only one type of finch present, if not extinct.

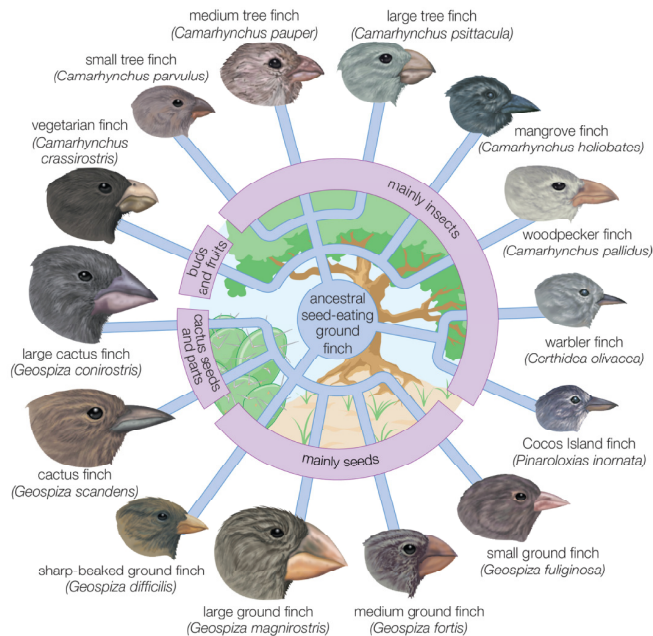


Figure 1.5. Adaptive radiation of Galapagos finches. The ancestral, seed-eating, ground finch diversified into a range of different species, when ecological opportunity was provided in the form of untapped resources. From [129].

1.3.2 Factors that influence adaptive radiation

Many studies have been performed that investigated environmental factors that influence the dynamics of AR. Examples of such factors include immigration history [125], disturbance of the environment [119], nutrient availability [130], competition and predation [131] or niche occupation [124]. However, as ecological opportunity is also influenced by the interaction between the founder of an AR and the environment, this means it is also influenced by the niche of the founder. Although some studies have been performed that examined the influence of founder characteristics on AR, relatively little is known. While comparative studies have revealed differences in the tempo and outcome of ARs that might be explained by different niches of the founders [132,133], phylogenetic reconstructions of extant ARs were less clear in their outcome [104,105]. In one experimental example of evolutionary branching, an effect of founder niche was observed and demonstrated that founder specialization can increase the likelihood of diversification [134].

In chapter 6 of this thesis, we used experimental evolution of the soil bacterium *Pseudomonas fluorescens* SBW25 as a model to examine the link between the evolvability and niche of organisms and their propensity for evolutionary diversification by AR. We accomplished this by decomposing organismal evolutionary constraints into two causative factors, namely evolvability and niche.

1.3.3 Model system: *Pseudomonas fluorescens* SBW25

In order to study adaptive radiation, we used experimental populations of the bacterium *Pseudomonas fluorescens* SBW25. Since its first publication in 1998 [118], this model system has been

used extensively to study several aspects of evolution, including bet hedging [135] and the influence of environmental factors [120] and immigration history [125] on adaptive radiation. When grown in statically incubated environments (microcosms) that contain nutrient-rich medium, *P. fluorescens* populations readily adapt to the spatially heterogeneous environment that consists of an oxygen-rich air-liquid interface and an oxygen-depleted liquid bulk. The ancestral genotype, which forms colonies of the smooth (SM) morphology class (Figure 1.6A) and grows predominantly as planktonic cells in static microcosms (Figure 1.6B), rapidly diversifies by mutation and selection into co-existing genotypes with different niches. These genotypes can be distinguished from the ancestral genotype both on the basis of their colony morphology and their growth-phenotype in the static microcosms. The colonies of these genotypes are of the wrinkly spreader (WS) class (Figure 1.6C) and in static microcosms they form thick biofilms at the air-liquid interface (Figure 1.6D) [118,136]. Overproduction of acetylated cellulose underlies the evolution of WS genotypes [136,137] and provides the opportunity for cells to stick together to colonize the air-liquid interface.

Previously, experiments showed that the SM-type ancestor and a WS genotype occupy different niches and that they are able to coexist due to negative frequency-dependent fitness interactions and trade-offs [106,118,136,138]. While both genotypes were capable of reciprocal invasion from rare in static microcosms, WS genotypes could not invade populations of the SM ancestor from rare when incubated in shaken microcosms. Similarly, differences between the niches of several WS genotypes were indicated, by showing that WS genotypes co-existed due to negative frequency-dependent fitness interactions [139]. The genetic changes underlying this model of adaptive radiation have been established for various niche specialists at the nucleotide level [135,136,140–144].

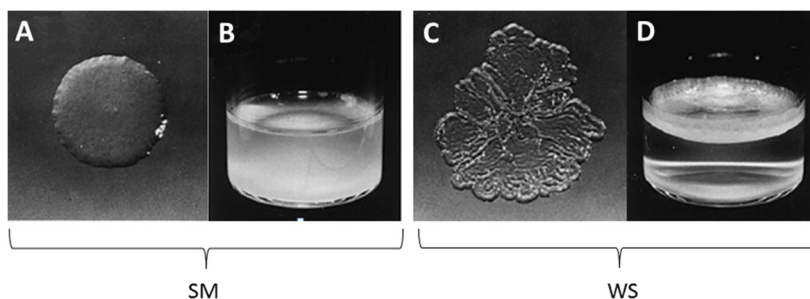


Figure 1.6. Phenotypic diversity in experimental populations of *Pseudomonas fluorescens* SBW25. A. SM colony morphology on agar plates. B. Niche of SM genotypes. C. WS colony morphology on agar plates. D. Niche of WS genotypes. Adapted from [118].

1.4 Thesis outline

In this thesis, we examined the evolution of protein complexes at the nanoscale and the evolution of ecological diversity at the macroscale. Although vastly different in terms of size scale, both studies contribute to our understanding of the evolution of biological complexity and diversity on Earth. In **Chapter 2–5**, we focused on the compositional evolution of protein complexes using the BFM of *E.*

coli. In **Chapter 6**, we examined how organismal properties can affect the evolution of biodiversity in an experimental adaptive radiation of the bacterium *Pseudomonas fluorescens*.

In **Chapter 2** we systematically investigated the creative potential of compositional evolution in the origin and evolution of protein complexes. We replaced the stator genes of *E. coli* with 22 different stator genes from 15 different species of bacteria and investigated for these 22 different strains both immediate compatibility of the foreign stators and the possibility for evolutionary integration. Our strategy successfully captured the evolutionary dynamics of foreign component integration and we identified several structural and functional characteristics of the foreign stators that possibly governed this. Furthermore, we showed that compositional evolution could lead to phenotypic variation and that different foreign stators resulted in different evolutionary dynamics. In all strains diverse, multi-step mutational trajectories were identified, that were in part a result of epistatic interactions with earlier mutations.

Chapter 3 focussed on the mutations underlying foreign component integration. We performed whole-genome re-sequencing and investigated predicted mutational differences both within and between strains. Furthermore, we focussed on the distribution of these mutations. Overall, we found that mutations were clustered, especially around the foreign stator proteins. Lastly, we identified both within and between strain differences in mutations, which indicated that there were several mutational trajectories to integrate a foreign component.

We examined the diversity in phenotypic and mutational trajectories underlying foreign component integration in **Chapter 4**, by directly investigating individual evolutionary lineages. We investigated underlying changes at lower-level, motility-related phenotypes and resolved the order of mutations in BFM genes and chemotaxis-signalling-pathway genes. Results showed that foreign component integration occurred through different, multi-step adaptive trajectories both within and between strains. Similarly, the pattern of evolutionary change in lower-level motility related phenotypes varied within and between strains.

The contribution of single mutations to functional integration of foreign components, was investigated in **Chapter 5**. We focussed on the first-step mutations in the foreign components that resulted in compatibility or improved integration. Within strain mutations indicated diversity in accessible mutations, but also identical mutations were found. Strain-specific constraints were identified by the diversity in mutations between different strains. Furthermore, we found that different mutations resulted in different motility levels and that mutations were not restricted to a specific part of the foreign component.

Finally, in **Chapter 6** we shifted our attention to the evolution of ecological diversity and examined organismal characteristics that shape adaptive radiation, which is the rapid diversification of a lineage into organisms with different niches. First, we made a distinction between organismal evolvability, which is the capacity of organisms to generate heritable and selectable phenotypic variation, and niche, which is the complex of reciprocal ecological interactions between organisms and the environment. Next, we examined the state of both characteristics in genotypes that had a reduced propensity for adaptive radiation. We found that genotypes with unchanged evolvability diversified less due to changes in their niches. With this result, we established the importance of the niche of the founder of adaptive radiations in relation to its outcome.

1.5 Literature

- 1 Darwin E (1794) *Zoonomia; or, the laws of organic life* J. Johnson, London.
- 2 Hutton J (1794) *An investigation of the principles of knowledge and of the progress reason, from sense to science and philosophy* Strahan & Cadell, Edinburgh.
- 3 Darwin C (1859) *On the origin of species by means of natural selection or the preservation of favoured races in the struggle for life* John Murray, London.
- 4 Mendel G (1866) Versuche über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereins Brünn* **4**, 3–47.
- 5 Huxley J (1942) *Evolution: the modern synthesis* Allen & Unwin, London.
- 6 Johanssen W (1909) *Elemente der exakten Erblchkeitslehre* Gustav Fisher, Jena.
- 7 Beadle G & Tatum E (1941) Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl. Acad. Sci.* **27**, 499–506.
- 8 Hershey A & Chase M (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* **36**, 39–56.
- 9 Dahm R (2008) Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.* **122**, 565–81.
- 10 Levene PA (1919) The structure of yeast nucleic acid: IV. Ammonia hydrolysis. *J. Biol. Chem.* **40**, 415–424.
- 11 Chargaff E (1950) Chemical Specificity of Nucleic Acids and Mechanism of their Enzymatic Degradation. *Experientia* **6**, 201–209.
- 12 Masel J (2011) Genetic drift. *Curr. Biol.* **21**, R837–8.
- 13 Whitlock M & Phillips P (1995) Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Syst.* **26**, 601–629.
- 14 Poelwijk FJ, Kiviet DJ, Weinreich DM & Tans SJ (2007) Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383–6.
- 15 Gabaldón T, Rainey D & Huynen M (2005) Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I). *J. Mol. Biol.* **348**, 857–70.
- 16 Forgac M (2007) Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. *Nat. Rev. Mol. Cell Biol.* **8**, 917–29.
- 17 Finnigan GC, Hanson-Smith V, Stevens TH & Thornton JW (2012) Evolution of increased complexity in a molecular machine. *Nature* **481**, 360–4.
- 18 Sowa Y & Berry RM (2008) Bacterial flagellar motor. *Q. Rev. Biophys.* **41**, 103–32.
- 19 Pallen M & Matzke N (2006) From The Origin of Species to the origin of bacterial flagella. *Nat. Rev. Microbiol.* **4**, 784–790.
- 20 Liu R & Ochman H (2007) Stepwise formation of the bacterial flagellar system. *Proc. Natl. Acad. Sci.* **104**, 7116–7121.
- 21 Mulkidjanian A & Makarova K (2007) Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. *Nat. Rev. Microbiol.* **5**, 892–899.
- 22 Dolezal P, Likic V, Tachezy J & Lithgow T (2006) Evolution of the molecular machines for protein import into mitochondria. *Science* **313**, 314–8.

- 23 Clements A, Bursac D, Gatsos X, Perry AJ, Civciristov S, Celik N, Likic VA, Poggio S, Jacobs-Wagner C, Strugnell RA & Lithgow T (2009) The reducible complexity of a mitochondrial molecular machine. *Proc. Natl. Acad. Sci.* **106**, 15791–15795.
- 24 Archibald JM, Logsdon Jr. JM & Doolittle WF (2000) Origin and Evolution of Eukaryotic Chaperonins: Phylogenetic Evidence for Ancient Duplications in CCT Genes. *Mol. Biol. Evol.* **17**, 1456–1466.
- 25 Seidl MF & Schultz J (2009) Evolutionary flexibility of protein complexes. *BMC Evol. Biol.* **9**, 155.
- 26 Rorick M (2012) Quantifying protein modularity and evolvability: a comparison of different techniques. *Biosystems.* **110**, 22–33.
- 27 Watson RA & Pollack JB (2005) Modular Interdependency in Complex Dynamical Systems. *Artif. Life* **11**, 445–457.
- 28 Gray MW, Lukeš J, Archibald JM, Keeling PJ & Doolittle WF (2010) Irremediable Complexity? *Science* **330**, 920–921.
- 29 Wagner GP, Pavlicev M & Cheverud JM (2007) The road to modularity. *Nat. Rev. Genet.* **8**, 921–31.
- 30 Gophna U & Ofra Y (2011) Lateral acquisition of genes is affected by the friendliness of their products. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 343–8.
- 31 Behe MJ (2001) Reply to My Critics: A Response to Reviews of Darwin's Black Box : The Biochemical Challenge to Evolution. *Biol. Philos.* **16**, 685–709.
- 32 Weinreich DM, Delaney NF, Depristo MA & Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111 – 114.
- 33 Watson RA (2002) *Compositional Evolution: Interdisciplinary Investigations in Evolvability, Modularity and Symbiosis.*
- 34 Paulick A, Koerd A, Lassak J, Huntley S, Wilms I, Narberhaus F & Thormann KM (2009) Two different stator systems drive a single polar flagellum in *Shewanella oneidensis* MR-1. *Mol. Microbiol.* **71**, 836–50.
- 35 Boto L (2010) Horizontal gene transfer in evolution: facts and challenges. *Proc. Biol. Sci.* **277**, 819–27.
- 36 Medrano-Soto A, Moreno-Hagelsieb G, Vinuesa P, Christen JA & Collado-Vides J (2004) Successful lateral transfer requires codon usage compatibility between foreign genes and recipient genomes. *Mol. Biol. Evol.* **21**, 1884–94.
- 37 Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P & Rubin EM (2007) Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**, 1449–52.
- 38 Delalez N & Armitage JP (2009) Parts exchange: tuning the flagellar motor to fit the conditions. *Mol. Microbiol.* **71**, 807–10.
- 39 Lind PA, Tobin C, Berg OG, Kurland CG & Andersson DI (2010) Compensatory gene amplification restores fitness after inter-species gene replacements. *Mol. Microbiol.* **75**, 1078–89.
- 40 Treangen TJ & Rocha EPC (2011) Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* **7**, e1001284.
- 41 Cohen O, Gophna U & Pupko T (2011) The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol. Biol. Evol.* **28**, 1481–9.
- 42 Omer S, Kovacs A, Mazor Y & Gophna U (2010) Integration of a foreign gene into a native complex does not impair fitness in an experimental model of lateral gene transfer. *Mol. Biol. Evol.* **27**, 2441–5.
- 43 Francino MP (2005) An adaptive radiation model for the origin of new gene functions. *Nat Genet* **37**, 573–578.
- 44 Kugelberg E, Kofoid E, Reams AB, Andersson DI & Roth JR (2006) Multiple pathways of selected gene amplification

- during adaptive mutation. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 17319–24.
- 45 Andersson DI & Hughes D (2009) Gene Amplification and Adaptive Evolution in Bacteria. *Annu. Rev. Genet.* **43**, 167–195.
- 46 Perica T, Chothia C & Teichmann S a (2012) Evolution of oligomeric state through geometric coupling of protein interfaces. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8127–32.
- 47 Hashimoto K & Panchenko AR (2010) Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20352–20357.
- 48 Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. sixth Int. Congr. Genet.* **1**, 356–366.
- 49 Weinreich DM, Watson R a & Chao L (2005) Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution (N. Y.)* **59**, 1165–74.
- 50 de Visser JAGM & Krug J (2014) Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* **15**, 480–490.
- 51 Berg HC (2003) The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19–54.
- 52 Blair DF (2003) Flagellar movement driven by proton translocation. *FEBS Lett.* **545**, 86–95.
- 53 Chen S, Beeby M, Murphy GE, Leadbetter JR, Hendrixson DR, Briegel A, Li Z, Shi J, Tocheva EI, Müller A, Dobro MJ & Jensen GJ (2011) Structural diversity of bacterial flagellar motors. *EMBO J.* **30**, 2972–81.
- 54 Larsen SH, Adler J, Gargus JJ & Hogg RW (1974) Chemomechanical Coupling without ATP: The Source of Energy for Motility and Chemotaxis in Bacteria. *Proc. Natl. Acad. Sci.* **71**, 1239–1243.
- 55 Manson MD, Tedesco PAT, Berg HC, Haroldt FM & van der Drift C (1977) A protonmotive force drives bacterial flagella. *Proc. Natl. Acad. Sci.* **74**, 3060–3064.
- 56 Zhou J, Sharp LL, Tang HL, Lloyd SA, Braun TF & Blair DF (1998) Function of Protonatable Residues in the Flagellar Motor of *Escherichia coli*: a Critical Role for Asp 32 of MotB. *J. Bacteriol.* **180**, 2729–2735.
- 57 Hirota N & Imae Y (1983) Na⁺-driven Flagellar Motors of an Alkalophilic *Bacillus* Strain YN-1. *J. Biol. Chem.* **258**, 10577–10581.
- 58 Braun TF, Poulson S, Gully JB, Empey JC, Van Way S, Putnam a & Blair DF (1999) Function of proline residues of MotA in torque generation by the flagellar motor of *Escherichia coli*. *J. Bacteriol.* **181**, 3542–51.
- 59 Hosking ER, Vogt C, Bakker EP & Manson MD (2006) The *Escherichia coli* MotAB proton channel unplugged. *J. Mol. Biol.* **364**, 921–37.
- 60 Chevance FF V & Hughes KT (2008) Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.* **6**, 455–65.
- 61 O'Brien E & Bennet P (1972) Structure of Straight Flagella from a Mutant *Salmonella*. *J. Mol. Biol.* **70**, 133–152.
- 62 Berg HC & Anderson RA (1973) Bacteria Swim by Rotating their Flagellar Filaments. *Nature* **245**, 380–382.
- 63 Larsen SH, Reader RW, Kort EN, Tso W-W & Adler J (1974) Change in direction of flagellar rotation is the basis of the chemotactic response in *Escherichia coli*. *Nature* **249**, 74–77.
- 64 Samatey FA, Matsunami H, Imada K, Nagashima S, Shaikh TR, Thomas DR, Chen JZ, DeRosier DJ, Kitao A & Namba K

(2004) Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism. *Nature* **431**, 1062–1068.

- 65 Depamphilis ML & Adler J (1971) Fine Structure and Isolation of the Hook-Basal Body Complex of Flagella from *Escherichia coli* and *Bacillus subtilis*. *J. Bacteriol.* **105**, 384–395.
- 66 Francis NR, Sosinsky GE, Thomas D & DeRosier DJ (1994) Isolation, Characterization and Structure of Bacterial Flagellar Motors Containing the Switch Complex. *J. Mol. Biol.* **235**, 1261–1270.
- 67 Katayama E, Shiraishi T, Oosawa K, Baba N & Aizawa S (1996) Geometry of the Flagellar Motor in the Cytoplasmic Membrane of *Salmonella typhimurium* as Determined by Stereo-photogrammetry of Quick-freeze Deep-etch Replica Images. *J. Mol. Biol.* **255**, 458–475.
- 68 Yamaguchi S, Aizawa S, Kihara MAY, Isomura M, Jones CJ & Macnab RM (1986) Genetic Evidence for a Switching and Energy-Transducing Complex in the Flagellar Motor of *Salmonella typhimurium*. *J. Bacteriol.* **168**, 1172–1179.
- 69 Welch M, Oosawat K & Aizawat S (1993) Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. *Proc. Natl. Acad. Sci.* **90**, 8787–8791.
- 70 Chun S & Parkinson J (1988) Bacterial motility: membrane topology of the *Escherichia coli* MotB protein. *Science* **239**, 276–278.
- 71 Khan S, Dapice M & Reese TS (1988) Effects of mot Gene Expression on the Structure of the Flagellar Motor at the Marine Biological Laboratory. *J. Mol. Biol.* **202**, 575–584.
- 72 Sarkar MK, Paul K & Blair D (2010) Chemotaxis signaling protein CheY binds to the rotor protein FliN to control the direction of flagellar rotation in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 9370–5.
- 73 Wadhams GH & Armitage JP (2004) Making sense of it all: bacterial chemotaxis. *Nat. Rev. Mol. Cell Biol.* **5**, 1024–37.
- 74 Zhou J & Blair DF (1997) Residues of the Cytoplasmic Domain of MotA Essential for Torque Generation in the Bacterial Flagellar Motor. *J. Mol. Biol.* **273**, 428–439.
- 75 Lloyd SA & Blair DF (1997) Charged Residues of the Rotor Protein FliG Essential for Torque Generation in the Flagellar Motor of *Escherichia coli*. *J. Mol. Biol.* **266**, 733–744.
- 76 Zhou J, Lloyd SA & Blair DF (1998) Electrostatic interactions between rotor and stator in the bacterial flagellar motor. *Proc. Natl. Acad. Sci.* **95**, 6436–6441.
- 77 de Mot R & Vanderleyden J (1994) The C-terminal sequence conservation between OmpA-related outer membrane proteins and MotB suggests a common function in both Gram-positive and Gram-negative bacteria, possibly in the interaction of these domains with peptidoglycan. *Mol. Microbiol.* **12**, 333–334.
- 78 Koebnik R (1995) Proposal for a peptidoglycan-associating alpha-helical motif in the C-terminal regions of some bacterial cell-surface proteins. *Mol. Microbiol.* **16**, 1269–1270.
- 79 Reid SW, Leake MC, Chandler JH, Lo C, Armitage JP & Berry RM (2006) The maximum number of torque-generating units in the flagellar motor of *Escherichia coli* is at least 11. *Proc. Natl. Acad. Sci.* **103**, 8066–8071.
- 80 Leake MC, Chandler JH, Wadhams GH, Bai F, Berry RM & Armitage JP (2006) Stoichiometry and turnover in single, functioning membrane protein complexes. *Nature* **443**, 355–8.
- 81 Braun TF, Al-mawsawi LQ, Kojima S & Blair DF (2004) Arrangement of Core Membrane Segments in the MotA / MotB Proton-Channel Complex of *Escherichia coli*. *Biochemistry* **43**, 35–45.
- 82 Kojima S & Blair DF (2004) Solubilization and Purification of the MotA / MotB Complex of *Escherichia coli*.

Biochemistry **43**, 26–34.

- 83 Xing J, Bai F, Berry R & Oster G (2006) Torque – speed relationship of the bacterial flagellar motor. **103**.
- 84 Adler J (1966) Chemotaxis in bacteria. *Science* **153**, 708–716.
- 85 Adler J (1966) Effect of amino acids and oxygen on chemotaxis in *Escherichia coli*. *J. Bacteriol.* **92**, 121–129.
- 86 Berg HC & Brown DA (1972) Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature* **239**, 500–504.
- 87 Blair DF & Berg HC (1990) The MotA protein of *E. coli* is a proton-conducting component of the flagellar motor. *Cell* **60**, 439–49.
- 88 Blair DF, Kim DY & Berg HC (1991) Mutant MotB Proteins in *Escherichia coli*. *J. Bacteriol.* **173**.
- 89 Che Y-S, Nakamura S, Kojima S, Kami-ike N, Namba K & Minamino T (2008) Suppressor analysis of the MotB (D33E) mutation to probe bacterial flagellar motor dynamics coupled with proton translocation. *J. Bacteriol.* **190**, 6660–7.
- 90 Wei Y, Wang X, Liu J, Nememan I, Singh AH, Weiss H & Levin BR (2011) The population dynamics of bacteria in physically structured habitats and the adaptive virtue of random motility. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4047–52.
- 91 Bren A & Eisenbach M (2000) How Signals Are Heard during Bacterial Chemotaxis : Protein-Protein Interactions in Sensory Signal Propagation. *J. Bacteriol.* **182**, 6865–6873.
- 92 Maddock JR & Shapiro L (1993) Polar Location of the Chemoreceptor Complex in the *Escherichia coli* Cell. *Science* **259**, 1717–1723.
- 93 Martin AC, Wadhams GH & Armitage JP (2001) The roles of the multiple CheW and CheA homologues in chemotaxis and in chemoreceptor localization in *Rhodobacter sphaeroides*. *Mol. Microbiol.* **40**, 1261–1272.
- 94 Hess JF, Oosawa K, Kaplan N & Simon MI (1988) Phosphorylation of three proteins in the signaling pathway of bacterial chemotaxis. *Cell* **53**, 79–87.
- 95 Toker A & Macnab R (1997) Distinct regions of bacterial flagellar switch protein FliM interact with FliG, FliN and CheY. *J. Mol. Biol.* **273**, 623–34.
- 96 Anand GS, Goudreau PN & Stock AM (1998) Activation of Methyltransferase CheB: Evidence of a Dual Role for the Regulatory Domain. *Biochemistry* **37**, 14038–14047.
- 97 Springer WR & Koshland DE (1977) Identification of a protein methyltransferase as the cheR gene product in the bacterial sensing system. *Proc. Natl. Acad. Sci.* **74**, 533–537.
- 98 McEvoy M, Bren A, Eisenbach M & Dahlquist F (1999) Identification of the binding interfaces on CheY for two of its targets, the phosphatase CheZ and the flagellar switch protein FliM. *J. Mol. Biol.* **289**, 1423–33.
- 99 Turner L, Ryu WS & Berg HC (2000) Real-Time Imaging of Fluorescent Flagellar Filaments Real-Time Imaging of Fluorescent Flagellar Filaments. *J. Bacteriol.* **182**, 2793–2801.
- 100 Lauffenburger DA (1991) Quantitative studies of bacterial chemotaxis and microbial population dynamics. *Microb. Ecol.* **22**, 175–85.
- 101 Osborn HF (1902) The Law of Adaptive Radiation. *Am. Nat.* **36**, 353–363.

- 102 Simpson G (1953) *The Major Features of Evolution* (Columbia Univ Press, New York).
- 103 Losos JB (2010) Adaptive radiation, ecological opportunity, and evolutionary determinism. *Am. Nat.* **175**, 623–39.
- 104 Glor RE (2010) Phylogenetic Insights on Adaptive Radiation. *Annu. Rev. Ecol. Evol. Syst.* **41**, 251–270.
- 105 Schluter D (2000) *The Ecology of Adaptive Radiation* Oxford University Press, Oxford.
- 106 Kassen R (2009) Toward a general theory of adaptive radiation: insights from microbial experimental evolution. *Ann. N. Y. Acad. Sci.* **1168**, 3–22.
- 107 Grant PR & Grant BR (2002) Adaptive radiation of Darwin's finches: Recent data help explain how this famous group of Galapagos birds evolved, although gaps in our understanding remain. *Am. Sci.* **130**, 130.
- 108 McInerney GJ & Etienne RS (2012) Pitch the niche - taking responsibility for the concepts we use in ecology and species distribution modelling. *J. Biogeogr.* **39**, 2112–2118.
- 109 McInerney GJ & Etienne RS (2012) Stitch the niche - a practical philosophy and visual schematic for the niche concept. *J. Biogeogr.* **39**, 2103–2111.
- 110 McInerney GJ & Etienne RS (2012) Ditch the niche - is the niche a useful concept in ecology or species distribution modelling? *J. Biogeogr.* **39**, 2096–2102.
- 111 Wagner CE, Harmon LJ & Seehausen O (2012) Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* **487**, 366–9.
- 112 Gillespie RG & Clague D (2009) *Encyclopedia of Islands* (Univ of California Press, Berkeley, CA).
- 113 Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. *Proc. Biol. Sci.* **273**, 1987–98.
- 114 Grant PR & Grant BR (2011) *How and Why Species Multiply* (Princeton Univ Press, Princeton).
- 115 Losos JB (2009) *Lizards in an Evolutionary Tree* (Univ of California Press, Berkeley, CA).
- 116 Helling R, Vargas CN & Adams J (1987) Evolution of *Escherichia coli* during growth in a constant environment. *Genetics* **358**, 349–358.
- 117 Rozen DE & Lenski RE (2000) Long-term experimental evolution in *Escherichia coli* VIII. Dynamics of a Balanced Polymorphism. *Am. Nat.* **155**, 24–35.
- 118 Rainey PB & Travisano M (1998) Adaptive radiation in a heterogeneous environment. *Nature* **32**, 69–72.
- 119 Buckling a, Kassen R, Bell G & Rainey PB (2000) Disturbance and diversity in experimental microcosms. *Nature* **408**, 961–4.
- 120 Kassen R, Llewellyn M & Rainey PB (2004) Ecological constraints on diversification in a model adaptive radiation. *Nature* **431**, 984–989.
- 121 Maharjan R, Seeto S, Notley-McRobb L & Ferenci T (2006) Clonal adaptive radiation in a constant environment. *Science* **313**, 514–517.
- 122 Habets MGJL, Rozen DE, Hoekstra RF & de Visser JAGM (2006) The effect of population structure on the adaptive radiation of microbial populations evolving in spatially structured environments. *Ecol. Lett.* **9**, 1041–8.
- 123 Craig MacLean R, Dickson A & Bell G (2004) Resource competition and adaptive radiation in a microbial microcosm. *Ecol. Lett.* **8**, 38–46.

- 124 Brockhurst M a, Colegrave N, Hodgson DJ & Buckling A (2007) Niche occupation limits adaptive radiation in experimental microcosms. *PLoS One* **2**, e193.
- 125 Fukami T, Beaumont HJE, Zhang X-X & Rainey PB (2007) Immigration history controls diversification in experimental adaptive radiation. *Nature* **446**, 436–9.
- 126 Bailey SF, Dettman JR, Rainey PB & Kassen R (2013) Competition both drives and impedes diversification in a model adaptive radiation. *Proc. Biol. Sci.* **280**, 20131253.
- 127 Kirschner M & Gerhart J (1998) Evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8420–8427.
- 128 Colegrave N & Collins S (2008) Experimental evolution: experimental evolution and evolvability. *Heredity (Edinb)*. **100**, 464–70.
- 129 Encyclopaedia Britannica (2015) Adaptive radiation. .
- 130 Kassen R, Buckling a, Bell G & Rainey PB (2000) Diversity peaks at intermediate productivity in a laboratory microcosm. *Nature* **406**, 508–12.
- 131 Meyer JR & Kassen R (2007) The effects of competition and predation on diversification in a model adaptive radiation. *Nature* **446**, 432–5.
- 132 Price T, Lovette IJ, Bermingham E, Gibbs H & Richman A (2000) The imprint of history on communities of North American and Asian warblers. *Am. Nat.* **156**, 354–367.
- 133 Ord T (2012) Historical contingency and behavioural divergence in territorial *Anolis* lizards. *J. Evol. Biol.* **25**, 2047–2055.
- 134 Spencer CC, Tyerman J, Bertrand M & Doebeli M (2008) Adaptation increases the likelihood of diversification in an experimental bacterial lineage. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 1585–9.
- 135 Beaumont HJE, Gallie J, Kost C, Ferguson GC & Rainey PB (2009) Experimental evolution of bet hedging. *Nature* **462**, 90–3.
- 136 Spiers AJ, Kahn SG, Bohannon J, Travisano M & Rainey PB (2002) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics* **161**, 33–46.
- 137 Spiers AJ, Bohannon J, Gehrig SM & Rainey PB (2003) Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. *Mol. Microbiol.* **50**, 15–27.
- 138 Remold S (2012) Understanding specialism when the Jack of all trades can be the master of all. *Proc. Biol. Sci.* **279**, 4861–9.
- 139 Brockhurst M a, Hochberg ME, Bell T & Buckling A (2006) Character displacement promotes cooperation in bacterial biofilms. *Curr. Biol.* **16**, 2030–4.
- 140 Goymer P, Kahn SG, Malone JG, Gehrig SM, Spiers AJ & Rainey PB (2006) Adaptive Divergence in Experimental Populations of *Pseudomonas fluorescens*. II. Role of the GGDEF Regulator, WspR, in Evolution and Development of the Wrinkly Spreader Phenotype. *Genetics* **173**, 515–526.
- 141 McDonald MJ, Gehrig SM, Meintjes PL, Zhang X-X & Rainey PB (2009) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. IV. Genetic constraints guide evolutionary trajectories in a parallel adaptive radiation. *Genetics* **183**, 1041–53.

-
- 142 McDonald MJ, Cooper TF, Beaumont HJE & Rainey PB (2011) The distribution of fitness effects of new beneficial mutations in *Pseudomonas fluorescens*. *Biol. Lett.* **7**, 98–100.
- 143 Bantinaki E, Kassen R, Knight CG, Robinson Z, Spiers AJ & Rainey PB (2007) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. *Genetics* **176**, 441–53.
- 144 Ferguson GC, Bertels F & Rainey PB (2013) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. V. Insight into the niche specialist fuzzy spreader compels revision of the model *Pseudomonas* radiation. *Genetics* **195**, 1319–1335.

2 Experimental demonstration of evolutionary integration of orthologous components into a protein complex

Régis C.E. Flohr, Thierry K.S. Janssens, Marcel Langelaan, Erwin van Rijn, Hubertus J.E. Beaumont. *In preparation for publication.*

2.1 Abstract

Protein complexes are composed of multiple different subunits that interact to perform a specific function. Most of what we know about the evolution of protein complexes comes from comparative genomics studies, but this understanding does not illuminate questions about the molecular mechanisms by which it is facilitated and their creative potential. One mode of evolution that appears to have played a key role in the origin, innovation and adaptation of protein complexes is compositional evolution. Compositional evolution is defined as the evolutionary joining of pre-existing proteins into complexes that perform a joint function. One requirement for compositional evolution to be possible is that the genes and the proteins they encode (both referred to as *foreign components*) are either immediately compatible with the host complex and other cellular systems or can be evolutionarily integrated along selectively accessible mutational trajectories. In this chapter, we examine the capacity of evolution to overcome component incompatibilities by systematically investigating the compatibility of 22 foreign, orthologous BFM components and the capacity of evolution to integrate incompatible components. Our model is the bacterial flagellar motor (BFM) of *Escherichia coli*, which is the rotary motor that powers bacterial motility. This model allows easy assessment of the compatibility of foreign components by investigation of the motility of populations of cells in semi-solid agar. Our results reveal that this system readily captures evolutionary integration events and that the potential of compositional evolution was possibly limited by several structural and functional factors. Furthermore, we empirically demonstrate that evolutionary integration of a single foreign component could occur along parallel genetic and phenotypic trajectories with vastly different phenotypic dynamics.

2.2 Introduction

Protein complexes such as for example the NADH:ubiquinone oxidoreductase [1], the V-type ATPase [2,3] or the bacterial flagellar motor (BFM) [4] are composed of multiple protein parts that are to some degree functionally and structurally independent. Although experimental insight into the mechanisms by which these intricate protein complexes evolved is scarce, comparative genomics studies suggest that they gradually evolved by the addition of pre-existing protein parts to simpler assemblies [1,5–11]. This mode of evolution is called compositional evolution [12] and requires either immediate compatibility of the pre-existing genes and corresponding protein parts (in terms of expression and functionality, respectively) or the capacity of mutations to generate compatibility. In this thesis we will use the term *foreign component* to indicate both genes and corresponding proteins that are (to be) used for compositional evolution. A rare example of a study that provides direct empirical data on compositional evolution, is recent work which showed that complexity in V-type ATPases could have increased by fitness-neutral gene duplication and divergence of the duplicate genes [3]. Other experimental studies have investigated the mechanisms behind compositional evolution more indirectly. For instance, two studies examined the consequences of replacing genes that encoded protein-complex parts for homologous genes [13,14] and provided evidence for both negative and neutral fitness consequences of the addition of foreign components. Furthermore, experimental data on protein oligomerization revealed that both mutations within and outside protein interaction interfaces could have equally large effects on the evolution of protein complexes (e.g. [15,16]). Despite insight from above mentioned studies, more experimental insight into the creative potential and underlying mechanisms of compositional evolution is required to further establish this fundamental process, which is an integral part of our understanding of the origins of biological complexity.

The goal of the work described in chapters 2-5 of this thesis was to gain new experimental insight into compositional evolution of protein complexes, an area that remains largely unexplored. In order to construct a conceptual framework that allows one to think about the creative potential of compositional evolution to generate protein complexes, consider a situation in which both gene duplication (followed by divergence) and horizontal gene transfer (HGT) [17–20] can result in the co-option of any protein-encoding gene that is present in the biosphere at a given point in time. When this is the case, ultimately, the creative potential of compositional evolution is constrained by only three factors: (i) the functional diversity of foreign components present in the biosphere at a point in time, (ii) the subset of the foreign components that is capable of interacting functionally with the host system and nascent protein complex and (iii) the capacity of evolution to forge functional interactions between the remaining foreign components in the biosphere and the host system and nascent protein complex. Together, these factors divide the foreign components eligible for compositional evolution into three domains: (i) foreign components that are already compatible and can be modified and functionally integrated by evolution, (ii) foreign components that are incompatible but can be modified and functionally integrated by evolution and (iii) foreign components that are incompatible and cannot be modified and functionally integrated by evolution within a defined interval of evolutionary time. Thus, in this view, the creative potential of compositional evolution is bound by what may be considered a *compatibility horizon* from which beyond no pre-existing foreign components can be co-opted for the origin, innovation and

adaptation of protein complexes. Where this horizon lies for a given complex with a certain (future) function depends on the fraction of pre-existing foreign components that are compatible or can be evolutionarily modified to be compatible.

The work described in this and the following chapters explores this conceptual compatibility horizon and the evolutionary dynamics that take place in its vicinity by using a model protein-complex. In this chapter, we develop our model system and begin to address the following questions. What are the predictors of compatibility of foreign components? What range of heritable phenotypic variation can be accessed by compositional evolution? What are the phenotypic dynamics of foreign component integration (are multi-step evolutionary trajectories possible, are alternative evolutionary trajectories possible, what is the influence of epistasis and can we identify differences between different strains) and can we say something about the underlying fitness landscapes?

The model system that we used to study the evolution of protein complexes is the bacterial flagellar motor (BFM) of *Escherichia coli* (*E. coli*). This extensively studied molecular machine (for reviews see [4,21,22]) is a rotary motor that resides in the bacterial cell wall and facilitates bacterial motility. It is composed of more than 20 different protein parts [21] and can be decomposed into four major structures: the rotating basal body, the stator complex, the flagellar filament and the hook that connects the basal body and flagellar filament [22]. Briefly, the *E. coli* BFM uses energy stored in the proton motive force (PMF) to generate torque and rotate the flagellum [23]. Ions flow through to the stator complex, bind to it transiently and flow into the cytoplasm [23], causing the complex to undergo conformational changes. The stator interacts with the basal body through electrostatic interactions [21,24–26]. As a result, the conformational changes in the stator complex cause the basal body to rotate. In turn, rotation of the basal body is translated via the hook and rod-proteins [4] to the flagellar filament. Several counter-clockwise (CCW) rotating filaments bundle together, forming one large, helical structure [27,28]. This bundle of flagella provides cells with a propeller that enables it to swim. Without additional control, BFM activity facilitates swimming in a random direction, which already confers an evolutionary selective advantage over non-motile cells by allowing cells to scavenge their environment for resources [29].

The selective advantage of motility is further increased by the ability of bacterial cells to control the direction of flagellar rotation in response to spatial gradients of chemicals in the process of chemotaxis [29–32]. This involves interaction of the BFM with the chemotaxis signalling pathway (see [33] for a review), which, for example, allows cells to undergo a biased random-walk in the direction of higher concentrations of a nutrient. Chemotactic swimming emerges as follows. During flagellar rotation, the direction can switch from counter clockwise (CCW) to clockwise (CW) for brief instances. The likelihood of such switching events can be controlled by the chemotaxis-signalling pathway in response to the binding of chemicals to the chemotaxis receptors. Temporary switching of the BFM from CCW to CW and back causes the swimming cells to randomly re-orient their swimming direction, so called tumbling. Tumbling is the result of disintegration of the flagellar bundle that is triggered by direction-switching of one of the filaments [34,35]. Thus, modulation of the tumbling frequency determines the duration of straight swimming (periods of straight swimming are called runs). When cells swim toward higher concentrations of nutrients, the tumbling frequency decreases and cells undergo prolonged runs. When cells swim toward lower concentrations of nutrients the tumbling frequency increases. This causes more frequent cell re-orientation and thus shorter runs. In concert,

suppression of tumbles when the cells experience increasing nutrient concentrations and their increase when decreasing concentrations are sensed, causes cells to move in the direction of higher concentrations of nutrients by a biased random walk.

The way we set out to systematically investigate compositional evolution was to focus on a key step in this process: functional integration of a foreign component. Because the details of these events are lost in evolutionary history, we approached them using an engineered system, which consisted of an *E. coli* cell with chimeric BFMs (see below), that was poised to evolve functional interactions with a non/sub-functional foreign component (*evolutionary integration* of the foreign component). With the goal of observing non-trivial evolutionary dynamics on an experimental-evolution time scale, we constructed a series of strains of *E. coli*, in which the stator genes had been replaced with homologous genes from other species across the bacterial kingdom. The homologous genes, 22 different homologues in total, were selected based on the availability of DNA sequences and additional organismal information in literature. The resulting BFMs were called *chimeric BFMs* (cBFMs) and the strains harbouring cBFMs were called *cBFM strains*. In total, we started our experiment with 22 cBFM strains. Because the stator complex, which is composed of two proteins (MotA and MotB [36,37]), is essential for BFM-mediated motility, this allowed us to follow evolutionary integration under selection for (improved) population motility. Although our experiment does not capture all integration events in the BFMs natural history, it does capture relevant instances [38] in a manner that is experimentally feasible.

We expected that our approach would yield cases in which a non- or sub-functional foreign component was integrated during experimental evolution (i.e. evolutionary integration). This was based on several lines of evidence. First, comparative genetics analyses have indicated that one of two stator complexes in *Shewanella oneidensis* MR-1 was acquired by HGT [38]. Second, previous work has demonstrated interspecies stator compatibility and the capacity of mutations to improve functionality (e.g. [39,40]). Third, multiple studies found compensatory mutations in BFM components following deleterious mutations in the same or other BFM components (e.g. [41–45]). Taken together, these findings indicate that our cBFM model system is amenable to experimental evolution.

The results described in this chapter showed that our approach was capable of capturing evolutionary dynamics of foreign component integration. The findings shed light on the phylogenetic range of foreign components that were compatible immediately and that of components that became compatible and/or became integrated by evolutionary modification. Several structural and functional factors that predicted this were identified. Compositional evolution resulted in heritable phenotypic variation in motility-related phenotypes, which included faster chemotactic swimming. The evolutionary phenotypic changes observed in replicate selection lines showed that integration of different foreign components occurred with radically different dynamics. These replicate selection lines also revealed the existence of different, multi-step trajectories in the stator-integration fitness landscape and indicated that these were, in part, caused by the epistatic effects of earlier mutations.

2.3 Results

We started with 10 replicate populations of cells (selection lines) that either harboured 1 of the 22 different sets of foreign stator genes (which resulted in cBFM strains) or harboured the wild-type

(WT) stator genes (reference strain). The stator genes were expressed from the low copy plasmid pBAD33 in all strains. The percentage amino acid identity of the foreign stator proteins relative to WT stator of *E. coli* ranged from 99.7% to 23.8%. The populations were founded on nutrient-rich, semi-solid, agar plates and incubated at 37°C, which allowed them to grow, and when the strain was able to swim, to spread out radially by chemotaxis toward the edge of the Petri dish. Every 24h (one selection round) all populations were checked for the generation of *flares*, which we defined as wedge-shaped populations of mutant cells that escaped from the edge of the colony at the site of inoculation or from the chemotaxis front of the ancestral population (Figure 2.3). The formation of a flare, which represented an increase in the ability to spread out in semi-solid agar, could be the result of an increase in growth rate when the ancestral population was non-motile, the emergence of motility when the ancestral population was non-motile or an increase in motility when the ancestral population was motile already. Because flare formation was a prerequisite under our selection regime for a population of cells to be directly transferred to the next selection plate, flare formation represented an increase in fitness (the ability to survive and reproduce) in our artificial environment. Alternatively, when no flares emerged, our regime selected for cells that managed to end up at the edge of the colony in the centre of the plate and thus selected for properties that facilitated that. This included an increase in growth rate that resulted in wedge-shaped colony sectors, similar to colony sectors that can be observed in colonies growing on top of agar [46–48].

Three different outcomes after a round of selection were possible. 1) If the ancestral population was motile and no flares were generated before it reached the edge of the semi-solid agar plate, then this marked the end of the selection line. 2) If the ancestral population was either motile and did not reach the edge of the semi-solid agar plate or non-motile and did not generate flares, then the population was incubated for another 24h. This was continued for up to one week (no checks during weekends), after which a sample of the population was taken. Part of this sample was transferred to fresh semi-solid agar for another round of selection (for up to four weeks including weekly transfer), while another part was stored at -80°C. 3) If evolution of (increased) motility had occurred (i.e. generation of one or more flares), then a sample of the flare was taken, from which a single genotype (i.e. cell) was selected and grown in batch culture. Part of the resulting population was used to inoculate a fresh semi-solid agar plate for an additional round of selection (for up to four weeks), while another part was stored at -80°C (see *Materials and Methods*). When such an evolved genotype reached the edge of the plate without the generation of flares, then that genotype was considered the end point of an evolutionary lineage.

In order to investigate whether random selection of a genotype from a flare resulted in the selection of a genotype that was different from the ancestor, we performed a preliminary screening of 50 genotypes isolated from a single flare. Results showed that all genotypes were more motile than their immediate ancestor, although phenotypic differences were observed (*Supplementary information* Figure 2.13 and Figure 2.14). Similar phenotypic differences were present between genotypes isolated from a population of cells harbouring WT stators (*Supplementary information* Figure 2.15). The underlying causes of the phenotypic differences were not investigated in more detail, and include genetic differences or differences in inoculum size.

2.3.1 Stator compatibility and functional integration

Table 2.1 shows key characteristics of the foreign stators (genes and/or corresponding proteins) and donor species that we used in this study. Our first objective was to investigate the compatibility of the foreign stators with the *E. coli* BFM and other cell systems and structures, by examining their ability to facilitate BFM-mediated motility on semi-solid agar. If a strain was motile directly upon insertion of the foreign stators, then the stators were considered compatible. If strains were non-motile, then the stators were considered incompatible. Several factors might underlie incompatibility, including stator-complex/BFM interaction, stator-complex/peptidoglycan interaction, ion selectivity of the stators, efficiency of membrane insertion of the stators and mRNA and protein expression levels.

We found 8 out of the 22 cBFM strains to be motile directly, although the resulting motility-phenotypes in semi-solid agar varied greatly (see section 2.3.2 and 2.3.3). Three of the directly-motile strains (strains 11, 17 and 18; compatibility group B) evolved improved motility in semi-solid agar during selection. The remaining five strains did not do so before they reached the edge of the plate (compatibility group A). Of the strains that harboured foreign stators that did not confer motility immediately (14 out of 22), 3 evolved compatibility between 6 and 27 days of selection (compatibility group C), whereas 11 strains did not evolve compatibility within 35 days of selection (compatibility group D), after which selection was ended.

The motility of all cBFM strains was also examined qualitatively by examining single cells under the microscope (data not shown). This demonstrated that all compatible stators also facilitated motility at the single-cell level, whereas none of the incompatible stators did so.

2.3.1.1 System properties predictive of compatibility and integration dynamics

Next we examined if characteristics listed in Table 2.1 correlated with initial compatibility of the foreign stators. In other words, we examined whether differences between groups A-B and C-D could explain why some stators conferred motility immediately, while others did not. Of the six characteristics tested, only the flagellation type of the donor species and the amino acid identity correlated with initial compatibility. The donor species of groups A and B were all peritrichously flagellated (not much is known about the location of flagella in the donor species of strain 26, except that it is non-motile [49]), whereas the donor species of groups C and D contained mostly lateral or polar flagella (Fisher exact probability test; $p < 0.001$). The difference between these types of flagellation is the positioning of flagella on the cell surface: peritrichous flagella are located all around the bacterial cell at random positions, lateral flagella are located only at the long sides of a cell and polar flagella are located at the cell poles. Although also the percentage amino acid identity differed significantly between groups A-B and C-D (Fisher exact probability test; classes 0-50% and 50-100%; $p = 0.002$), the strains of group B were significantly different from those in group A in this same respect (Fisher exact probability test; classes 0-50% and 50-100%; $p = 0.018$). This suggested that amino acid identity was not a reliable predictor for initial compatibility.

Next, we investigated if differences in the characteristics between compatibility groups A and B (listed in Table 2.1) could explain the observed differences in the repeatability and timing of the evolution of improved motility between these groups. The repeatability of evolution was defined as the number of replicate selection lines founded with a single strain that evolved (improved) motility

during selection. The timing of evolution was defined as the number of days a selection line required before the first flare was generated. Besides the percentage amino acid identity of the foreign stators, which as mentioned previously differed between groups A and B, we found that the phylogenetic class and gram-staining of the donor species also correlated with differences in the ability to evolve improved motility between groups A and B (Fisher exact probability test; $p = 0.018$ for both comparisons).

We found no significant differences in stator characteristics that explained the observed differences in the repeatability and timing of evolution between compatibility groups C and D (Fisher exact probability test; $p > 0.05$ for all characteristics).

2.3.2 Compatible stators that were not evolutionarily integrated

Of the 22 foreign stators, 23% (5 out of 22) was compatible immediately and resulted in strains that did not evolve an improved motility during selection (group A in Table 2.1). The donor species of these five stators belonged to the same phylogenetic class as our acceptor species (*E. coli* K-12 MG1655; γ -proteobacteria) and their amino acid identity relative to the concatenated WT stator proteins ranged from 99.7% to 56.7%. All five strains showed chemotaxis-driven motility similar to a strain with WT stators (Figure 2.3A) and no differences in growth rate were observed between these strains and the reference strain (data not shown). Because comparative genetic studies suggest that compositional evolution is an important mode for the generation of heritable, phenotypic variation for adaptive evolution of protein complexes [38], we set out to investigate the phenotypic effects of introduction of these foreign stator proteins. We did this by examining five phenotypes of the five cBFM strains with compatible stators that did not show evolutionary integration: population swimming speed (i.e. speed of the chemotactic population front), competitive motility (population-level colonisation success in competition with a fluorescently-labelled reference strain), single-cell swimming speed, cell surface (as a proxy for cell size) and the percentage non-motile cells in a population.

2.3.2.1 Relative population swimming speed

Because the selection experiment was performed in semi-solid agar, the population phenotypes *relative swimming speed* and *competitive motility* were also measured in semi-solid agar. As such, these phenotypes represented components of fitness that governed the evolutionary dynamics during our selection experiment.

First, we investigated the relative population swimming speed. This was achieved by measuring the distance travelled by the front of a population per hour (during the first 8h after inoculation) in semi-solid agar and relating this to the reference strain (Figure 2.1A). All strains of compatibility group A showed chemotaxis-driven motility similar to the reference strain, which is characterised by two rings of cells that radiate outward from the point of inoculation (Figure 2.3A) [30]. The first ring is formed by cells that perform chemotaxis towards the most favourable nutrient (serine in LB medium [30]) and use all oxygen [30]. The second ring is composed of cells that consume residuals of the first-utilised nutrient anaerobically or perform chemotaxis towards other nutrients, while using oxygen

Table 2.1. Key characteristics of the foreign stators and donor species used in this study. All stator genes were expressed from the low copy plasmid pBAD33. Strain 01 is the strain that harboured WT *E. coli* stators (reference strain). *Amino acid identity* is expressed as the percentage of identical amino acids between the foreign and wild-type stator proteins. *Repeatability and timing of evolution*: 10/10 indicates that 10 out of 10 replicate lines evolved compatibility or improved motility. 3-13d indicates that the generation of all first flares occurred between 3 and 13 days of selection. *Compatibility groups* are defined on the basis of a combination of *compatibility* and *repeatability* values.

Strain ID	Stator genes (ion selectivity)	Donor species	Phylogenetic class donor species	Gram-staining donor species	Flagellation type donor species	Length concatenated stator genes	Amino acid identity [%]	GC content [%]	Compatibility	Repeatability and timing of evolution	Compatibility group
01	<i>motAB</i> (H ⁺)	<i>Escherichia coli</i> MG1655 K-12	γ -proteo	-	peritrichous	603	100	52.27	Motile	No evolution	A
02	<i>motAB</i> (H ⁺)	<i>Escherichia coli</i> O111:H-str. 11128	γ -proteo	-	peritrichous	603	99.67	52.16	Motile	No evolution	
03	<i>motAB</i> (H ⁺)	<i>Escherichia coli</i> O26:H11 str. 11368	γ -proteo	-	peritrichous	603	99.66	52.16	Motile	No evolution	
22	<i>motAB</i> (H ⁺)	<i>Edwardsiella ictaluri</i> 93-146	γ -proteo	-	peritrichous	571	72.60	56.11	Motile	No evolution	
24	<i>motAB</i> (H ⁺)	<i>Proteus mirabilis</i> HI4320	γ -proteo	-	peritrichous	646	64.76	41.31	Motile	No evolution	
26	<i>motAB</i> (H ⁺)	<i>Sodalis glossinidius</i> str. 'morsitans'	γ -proteo	-	unclear, non-motile	639	56.65	52.43	Motile	No evolution	
11	<i>motPS</i> (Na ⁺)	<i>Bacillus pseudofirmus</i> OF4	Firmicutes	+	peritrichous	518	25.09	40.11	Motile	10/10; 3-13d	B
17	<i>motAB</i> (H ⁺)	<i>Listeria monocytogenes</i> EGD-e	Firmicutes	+	peritrichous	558	23.81	42.74	Motile	10/10; 2-3d	
18	<i>motPS</i> (Na ⁺)	<i>Bacillus megaterium</i> DSM319	Firmicutes	+	peritrichous	470	25.22	40.32	Motile	10/10; 2-6d	
08	<i>lafTU</i> (H ⁺)	<i>Escherichia coli</i> O111:H-str. 11128	γ -proteo	-	lateral	594	34.20	54.08	Non-motile	3/10; 6-7d	C
13	<i>lafTU</i> (H ⁺)	<i>Photobacterium profundum</i> SS9	γ -proteo	-	lateral	612	35.08	43.57	Non-motile	6/10; 13-27d	
29	<i>motAB</i> "1" (H ⁺)	<i>Rhodospirillum centenum</i> SW	α -proteo	-	lateral	619	38.88	67.47	Non-motile	9/10; 6-13d	
05	<i>motAB</i> "1" (H ⁺)	<i>Bdellovibrio bacteriovorus</i> HD100	δ -proteo	-	polar	514	30.86	50.15	Non-motile	No evolution	D
09	<i>motAB</i> (H ⁺)	<i>Azospirillum</i> sp. B510	α -proteo	-	lateral/polar	642	35.18	69.10	Non-motile	No evolution	
14	<i>lafTU</i> "2" (H ⁺)	<i>Vibrio shilonii</i> AK1	γ -proteo	-	lateral	610	31.63	43.45	Non-motile	No evolution	
15	<i>pomAB</i> (Na ⁺)	<i>Shewanella oneidensis</i> MR-1	γ -proteo	-	polar	563	27.42	45.84	Non-motile	No evolution	
16	<i>motAB</i> (H ⁺)	<i>Bacillus megaterium</i> DSM319	Firmicutes	+	peritrichous	541	25.27	39.16	Non-motile	No evolution	
19	<i>MotAB putative</i> (Na ⁺)	<i>Photobacterium profundum</i> SS9	γ -proteo	-	polar	563	25.26	42.31	Non-motile	No evolution	
20	<i>lafTU</i> "1" (H ⁺)	<i>Vibrio shilonii</i> AK1	γ -proteo	-	lateral	614	32.34	48.97	Non-motile	No evolution	
21	<i>pomAB</i> (Na ⁺)	<i>Vibrio shilonii</i> AK1	γ -proteo	-	polar	510	25.84	47.46	Non-motile	No evolution	
23	<i>motAB</i> (H ⁺)	<i>Shewanella oneidensis</i> MR-1	γ -proteo	-	polar	519	24.51	37.71	Non-motile	No evolution	
25	<i>motAB</i> (H ⁺)	<i>Pseudomonas aeruginosa</i> UCBBP-PA14	γ -proteo	-	polar	630	42.15	65.37	Non-motile	No evolution	
28	<i>motCD</i> (H ⁺)	<i>Pseudomonas aeruginosa</i> UCBBP-PA14	γ -proteo	-	polar	542	26.69	66.25	Non-motile	No evolution	

that has diffused into the upper layer of the agar [30,31,50]. The population of cells that lags behind has entered stationary phase [50]. The population swimming speed is the result of a combination of many underlying, microscopic factors, including single-cell swimming speed, tumbling frequency, chemotaxis signalling efficiency and attractant consumption rate (gradient formation rate).

Strains 22 and 26 were both significantly slower than the reference strain (Benjamini-Hochberg-corrected Student's *t*-test; $n=9$; $p<0.001$ for both comparisons). Interestingly, three of the strains harbouring foreign stators were significantly faster than strain 01 in this assay (Benjamini-Hochberg-corrected Student's *t*-test; $n=9$; $p<0.05$ for all comparisons), of which strain 24 was fastest.

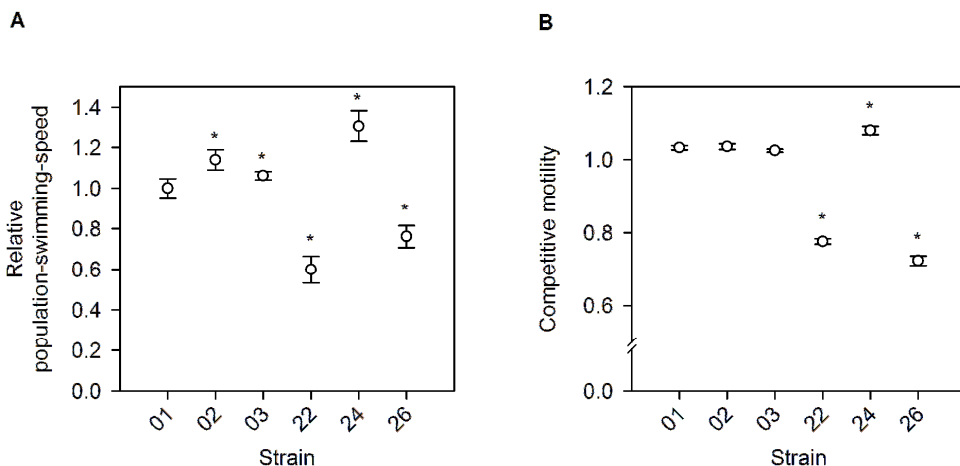


Figure 2.1. Fitness-related population phenotypes of strains harbouring compatible foreign stators. **A.** Population swimming speeds, relative to the reference strain harbouring WT *E. coli* stator genes (strain 01). $n = 9$. **B.** Competitive motility against a fluorescently-labelled version of the reference strain harbouring WT *E. coli* stator genes. Competitive motility is defined as the ratio of expansion velocities between competing strains (*Materials and Methods*). $n = 6$. Asterisks indicate significant deviation from strain 01. Bars indicate standard deviation. See text for statistics.

2.3.2.2 Competitive motility

The population swimming speeds were measured in isolation, which meant that only one strain was inoculated per semi-solid agar plate. Although this measurement provides meaningful insight into a fitness-related trait, it does not consider ecological interactions with other strains. However, such interactions are important during the selection experiment. Because the selection experiment favoured genotypes that were fitter on semi-solid agar in direct competition with their immediate ancestor, competition measurements provide a more accurate measure of the fitness effect of the introduction of the foreign stator genes in our experimental environment. In order to get insight into the fitness effects of foreign stators in comparison with WT stators, we also measured the ability of strains with foreign stators to compete with a fluorescently-labelled version of the reference strain (strain 01-mKate2; *Materials and Methods*). From these competitions, we determined the relative fitness by quantifying the outcome of the initial stage of competitive colonization of the semi-solid agar plate [48], which we termed *competitive motility*. As is shown in Figure 2.1B, strain 01 showed a higher competitive motility (1.036) than the fluorescent variant of the same strain. This was

expected due to the metabolic costs of expressing the fluorescent protein and will be of no influence on the results discussed below, as we will use the value of competitive motility of strain 01 as reference to investigate the fitness effects of all foreign stators.

Strain 22 and 26 both had a significantly lower competitive motility than the reference strain harbouring WT stators (Figure 2.1B; Benjamini-Hochberg-corrected Student's *t*-tests; $n=6$; $p<0.001$ for both comparisons). Although this is in agreement with the results for relative population swimming speed, the relative population swimming speed of strain 22 was significantly slower than that of strain 26 (Figure 2.1A), while its competitive motility was significantly higher than that of strain 26 (Benjamini-Hochberg-corrected Student's *t*-tests; $p<0.001$ for both comparisons). This discrepancy might be explained by ecological interactions between the competing genotypes, for instance, sensing of the nutrient gradient that is created by the other strain or chemical repulsion [51]. The difference that was observed between the relative population swimming speeds of strains 02 and 03 (in comparison with strain 01) was not present when we analyzed their competitive motility (Benjamini-Hochberg-corrected Student's *t*-tests; $n=6$; $p>0.05$ for both comparisons). These results might be explained by a lack of the ecological interactions mentioned above and show that the phenotypic effects of replacing the stator genes depends on both genetic and environmental factors (genotype-by-environment interaction; e.g. [52]). Strain 24 on the other hand, showed a significantly higher competitive motility compared with strain 01, similar to the result found for relative population swimming speed (Benjamini-Hochberg-corrected Student's *t*-test; $n=6$; $p<0.001$).

Taken together, the introduction of foreign stator genes resulted in some cases in heritable, phenotypic variation that can be favoured by natural selection. Because the macroscopic population-phenotypes that we described in this section result from changes in lower-level microscopic phenotypes, we next investigated a few of these microscopic phenotypes in order to provide deeper insight into the effects of expressing foreign stator genes.

2.3.2.3 Single-cell motility phenotypes

The strains that we investigated at the single-cell level were strain 01 (harbouring WT stators), strain 24 (faster than strain 01) and strain 26 (slower than strain 01). Our focus in this thesis was on three single-cell phenotypes that are important for motility in semi-solid agar, although other microscopic factors underlying motility exist (e.g. chemotactic efficiency, tumbling frequency, number of flagella and attractant consumption rate). First, we investigated the single-cell swimming speed in liquid and the percentage of non-motile cells in the population. While the percentage of non-motile cells might affect the swimming speed of the population by affecting nutrient-gradient formation, which is important for chemotaxis [53], the single-cell swimming speed might affect population swimming speed by affecting the efficiency of following the nutrient gradient (see *Discussion*). Furthermore, both phenotypes provide more direct insight into the functioning of the BFM. The third phenotype we examined is the cell surface (as a proxy for cell size), because it may influence the ability of cells to move through the pores of the semi-solid agar and thus influence the swimming speed of the population.

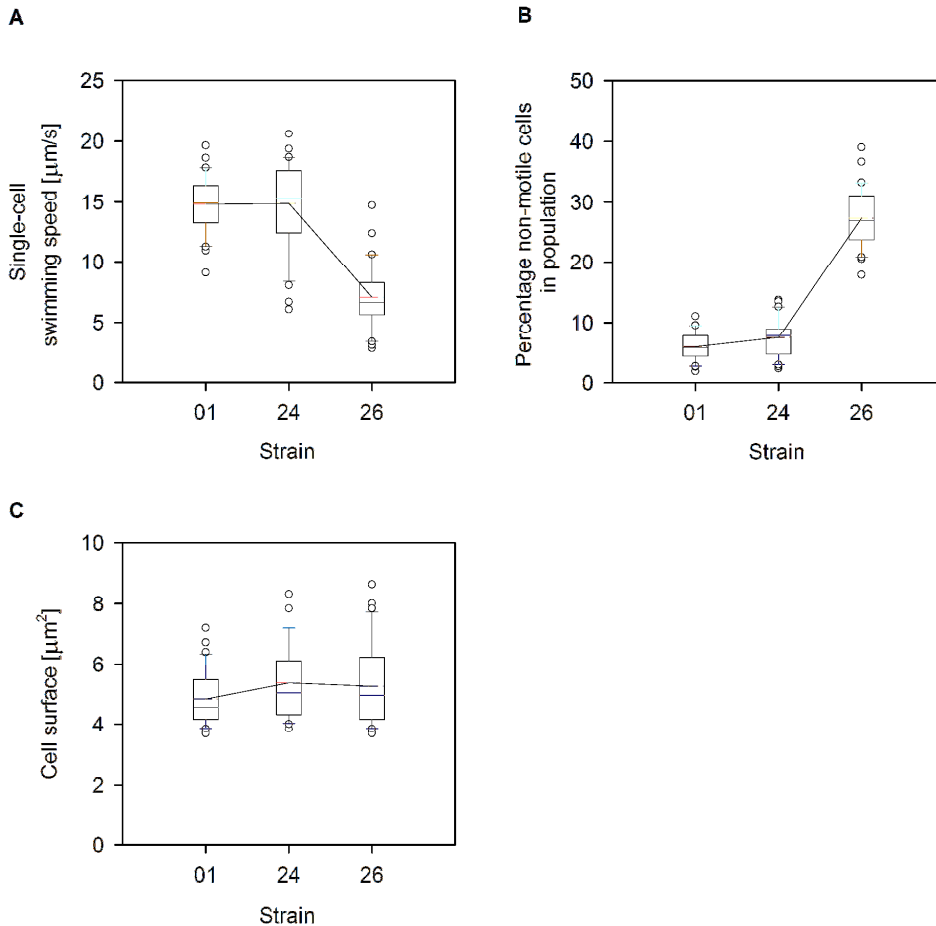


Figure 2.2. Single-cell phenotypes of two strains harbouring compatible foreign stators. **A.** Single-cell swimming speed measured in a liquid medium. **B.** The percentage of non-motile cells in the population. **C.** Cell surface. Asterisks indicate significant deviation from strain 01. $n = 30$. Boxes indicate the middle 50% of data points, with whiskers above and below indicating the 90th and 10th percentile, respectively. Circles indicate data points outside the 10 and 90th percentile. The horizontal lines in boxes that are interconnected by a line indicate means. The other horizontal line in a box indicates the median. See text for statistics.

Similar to our results at the population level, strain 26 was significantly slower at the single-cell level than a strain harbouring WT stators (Figure 2.2A; Benjamini-Hochberg-corrected Mann-Whitney U -test; $n=30$; $p<0.001$). In contrast to our results on semi-solid agar, strain 24 could not be distinguished from strain 01 at the single-cell level on the basis of its swimming speed (Benjamini-Hochberg-corrected Mann-Whitney U -test; $n=30$; $p=0.843$). The single-cell swimming speed of strain 01, which is dependent on several factors including temperature, pH and medium composition [4,54], was similar to speeds found in literature: approximately 10 body lengths per second ($\sim 20 \mu\text{m/s}$) [55].

The percentage of non-motile cells (Figure 2.2B) was significantly higher for strain 26 than it was for strain 01 (Benjamini-Hochberg-corrected Mann-Whitney U -test; $n=30$; $p<0.001$), while strain 24

showed a percentage of non-motile cells that was indistinguishable from WT (Benjamini-Hochberg-corrected Mann-Whitney U -test; $n=30$; $p=0.618$).

Cell surface measurements (Figure 2.2C) did not reveal any significant differences (Benjamini-Hochberg-corrected Mann-Whitney U -test; $n=30$; $p>0.05$ for both comparisons).

Next, we will examine the strains of compatibility group B, which harboured compatible stators and evolved improved motility upon selection (i.e. showed evolutionary integration). We will focus on the dynamics of evolution of improved motility and the within- and between-strain variability herein.

2.3.3 Evolutionary integration of compatible stators

Three of the cBFM strains that contained stators that were compatible immediately, evolved improved motility upon selection in semi-solid agar (group B in Table 2.1). The donor species of the stator genes in these strains are from the bacterial class of Firmicutes and are gram-positives in contrast to the *E. coli* host strain, which is gram-negative. The percentage of amino acid identity of the foreign stators relative to WT stators was around 25%, which is significantly lower than the percentage amino acid identity of the stators of group A (see section 2.3.1.1).

In contrast to the strains of compatibility group A, which showed WT motility (Figure 2.3A), strains 11, 17 and 18 displayed a different kind of motility. The phenotype was characterized by dots radiating from the point of inoculation (after 24h), which represented small bacterial colonies growing in the agar surrounding the inoculation site (Figure 2.3C), and was not observed for a non-motile, statorless strain (Figure 2.3B). We refer to this type of motility as atypical motility (ATM), which was previously reported in literature [25,56–58] to be the result of the presence of a few motile cells in the population that usually produce non-motile daughter cells (but see *Discussion*). Because strains performing ATM moved very slowly on semi-solid agar (approximately 4mm in 24h), no relative population swimming speeds could be determined.

When a mutation in a cell of a population, that showed ATM, resulted in an improved ability to spread out in semi-solid agar, this caused formation of a wedge-shaped flare escaping from the ancestral population (red ellipse in Figure 2.3D). In many cases, genotypes isolated from such a primary flare (for explanation of nomenclature see *Materials and Methods* Figure 2.10) evolved improved motility when transferred to fresh semi-solid agar. Examples of different flare phenotypes that were observed throughout the selection experiment for different (evolved) cBFM strains are shown in Figure 2.3E-G.

Below we will discuss the evolution of improved motility of strains that contained compatible stators in more detail. First, we will qualitatively describe the overall structure of the evolutionary ‘trees’ of (branching) replicate lineages that emerged during the selection experiments from the cBFM strains. Then we will quantitatively describe the relative population swimming speeds of the evolved genotypes at the end of each replicate evolutionary lineage. Both results provide insight into the repeatability of the evolution of improved motility and shed light on the structure of the fitness landscape that governs this, in terms of the availability of different adaptive trajectories.

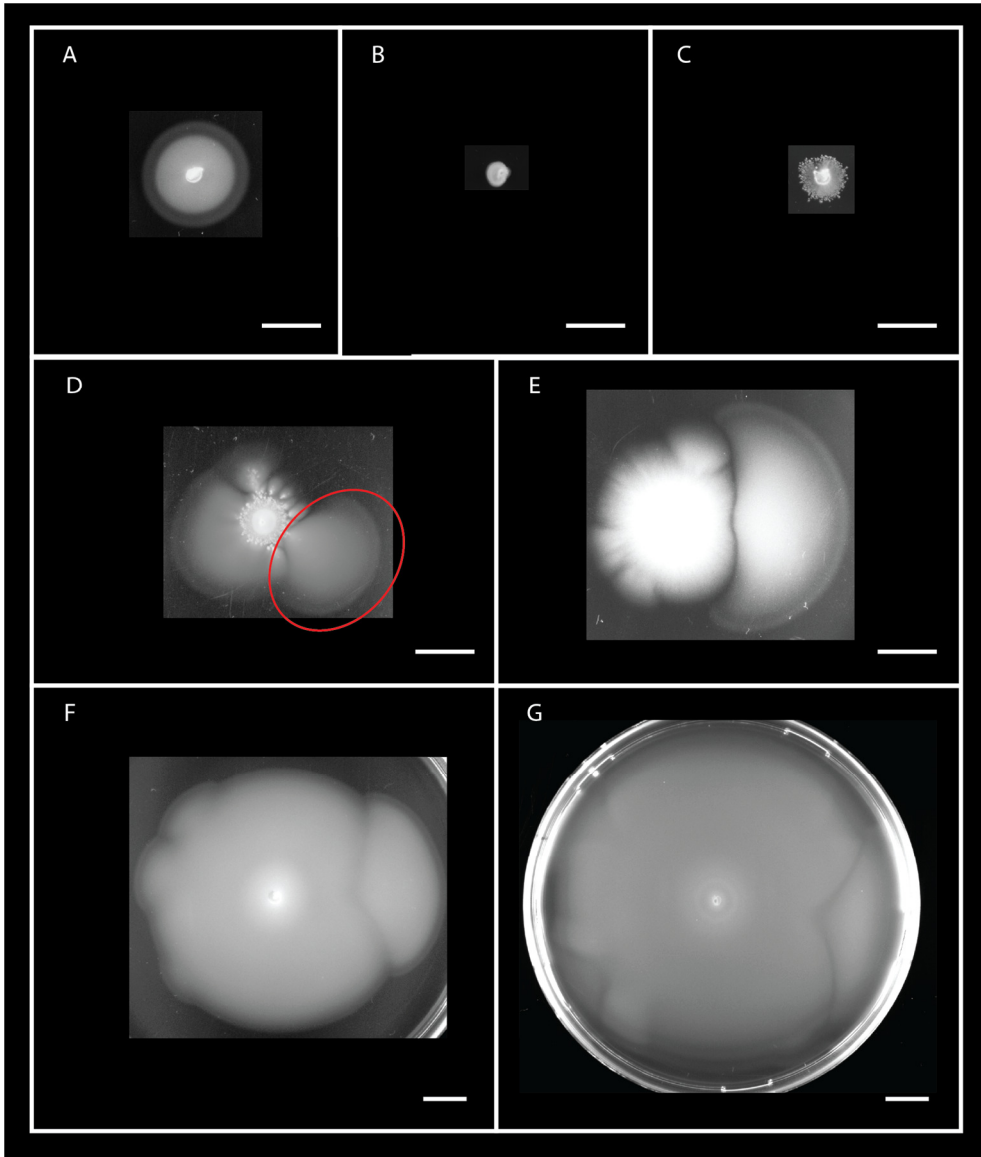


Figure 2.3. Different motility phenotypes of populations of cells in semi-solid agar. Flares indicate the evolution of stator-compatibility or improved motility in cBFM strains. **A.** Wild-type motility, characterised by several rings that radiate outward from the point of inoculation. The first ring is formed by cells that perform chemotaxis towards the most favourable nutrient (serine in LB medium [30] and use all oxygen. The second ring consumes residual nutrients anaerobically or performs chemotaxis towards other nutrients, while using oxygen that has diffused into the upper layer of the agar [30,31,50]. The population of cells that lags behind has entered stationary phase [50]. **B.** A non-motile, statorless strain which showed only growth. **C.** Atypical motility (ATM), characterised by dots radiating from the point of inoculation. **D.** Evolution of improved motility. A wedge-shaped flare (red ellipse) was generated by mutated cells from the ancestral population, which in this case showed ATM. **E-G.** Examples of additional flare phenotypes. All pictures were taken after approximately 24h. White bars indicate 1cm.

2.3.3.1 Strain 11

We started with 10 replicate populations of strain 11, which harboured a sodium stator from *Bacillus pseudofirmus* OF4 [59–61] as opposed to the proton stator that is normally present in *E. coli*. Nine populations evolved improved motility between 3 and 7 days of selection, while the 10th population required 13 days to evolve improved motility for the first time (yielding primary flare 11.1k).

Figure 2.4A shows the evolutionary lineages of strain 11 that were generated during the selection experiment. An evolutionary lineage is represented by a sequence of genotypes that were isolated from consecutive flares. For instance, genotype 11.1j was isolated from the primary flare and was the direct ancestor of the genotype that formed the secondary flare, from which 11.1j.2a was isolated. This genotype in turn gave rise to a tertiary flare, from which genotype 11.1j.2a.3b was isolated. The length and number of genotypes in each lineage was determined by the number of consecutive flares. In the previous example, the length of the lineage was three. Lineages branched when a population of a genotype gave rise to more than one flare. For instance, genotype 11.1j.2a gave rise to three tertiary flares and thus branched into three genotypes that were isolated from these flares. Letters were assigned to each flare after a branching point to make a distinction between flares, e.g. 11.1j.2a.3a, 11.1j.2a.3b, 11.1j.2a.3c. When a genotype isolated from a flare did not generate a flare upon transfer to fresh semi-solid agar and before it reached the edge of the plate, then the selection was ended and the last genotype was termed the endpoint of the lineage.

Almost all evolutionary lineages of strain 11 had a length of one or two flares. Only the lineage starting with genotype 11.1j yielded tertiary and quaternary flares. Although multiple mutations can occur before generation of a flare, in this chapter (sequencing was performed later; Chapter 3) we operate under the assumption that, overall, there is a correlation between the number of flares and the number of mutations in a lineage. Based on this, variation in trajectory length between the lineages indicated differences in the number of mutations required to reach an endpoint genotype.

To gain more insight into the diversity of genotypes that evolved during the selection experiment, and thus insight into the diversity of the underlying mutational trajectories, we measured the relative population swimming speeds of all endpoints that were sequenced (indicated in red in Figure 2.4A; see Chapter 3), as a measure of the ability to spread out in semi-solid agar. This ability was found to be a heritable phenotype, because genotypes selected from flares (that by definition evolved an improved ability to spread out in semi-solid agar) also showed an improved ability to spread out (observation). Following the currently accepted paradigm that phenotypes that breed true in bacteria are genetically encoded [62,63], this means that differences in relative population swimming speed must represent differences in genotype. Statistical analysis indicated significant differences in relative population swimming speeds between the evolved endpoints (Figure 2.4B; Benjamini-Hochberg-corrected Kruskal-Wallis test; $n = 3$, $p < 0.001$) and thus demonstrated differences in their mutational trajectories.

Next we investigated the correlation between the length of an evolutionary lineage and the relative population swimming speed of the corresponding endpoint, as this would shed light on the structure of the local fitness landscape and degree of epistasis that govern evolutionary integration of a foreign component. A negative correlation would indicate that the highest relative population swimming speed was reached, on average, by shorter lineages with one or a few mutational steps of

large fitness benefit (*big-benefit mutations*). Thus, when such a negative correlation between lineage length and endpoint fitness is detected with statistical significance, then this must involve a genetic bias in the longer lineages that either prevent the occurrence of a large positive effect of the mutations that confer a large beneficial effect in the shorter lineages (magnitude epistasis [64]), prevent the selection of the mutations due to negative fitness effects (sign epistasis [64]) or prevent the occurrence of the mutations altogether. In other words, these big-benefit mutations must have been less selectively-accessible after a certain point in the longer lineages. A possible mechanism behind this is epistatic interactions engendered by an early mutation, which constrain the possibilities for future big-benefit mutations. A positive correlation between the length of an evolutionary lineage and the relative population swimming speed of the corresponding endpoint indicates, on average, a lack of early mutations that confer a big benefit in terms of component integration and associated increase in fitness. This would thus indicate constraints due to epistatic interactions with the foreign stator. For strain 11, we did not find a significant correlation between lineage length and relative population swimming speed of endpoints (Figure 2.4C; Benjamini-Hochberg-corrected Spearman's correlation; $\rho = 0.046$; $p = 0.876$). This indicated that high motility could be reached with both few and many mutations and that epistasis was not very strong.

The data shown in Table 2.1 reveals that strains from compatibility group C required more time to generate primary flares in comparison with strains from compatibility group B. Inspired by this observation, we investigated the correlation between the rank of consecutive flares (primary, secondary, tertiary, etc.) and the time required for their generation (number of days of selection after inoculation). For instance, how many days were required to generate the primary flares from the ancestral strain and what was the number of days required to generate secondary flares from genotypes that were isolated from primary flares? This characteristic of evolution is determined by the incidence of mutations that increased motility and the degree to which they increased fitness (i.e. the distribution of fitness effects [65]). A shorter time to generate flares is associated with a higher availability of mutations, larger beneficial fitness effects or a combination thereof [66,67]. Accordingly, a negative correlation indicates that with increasing motility, more mutations became available that increased motility further and/or that these mutations conferred a larger fitness benefit. A positive correlation indicates that with increasing motility, fewer mutations that increased motility further were accessible and/or that their beneficial effects on fitness were smaller. Our results showed a significant, negative correlation for strain 11 (Figure 2.4D; Spearman's correlation; $\rho = -0.858$; $p < 0.001$), which thus indicated that more mutations to increase motility became available and/or that the fitness effect of mutations increased with increasing motility.

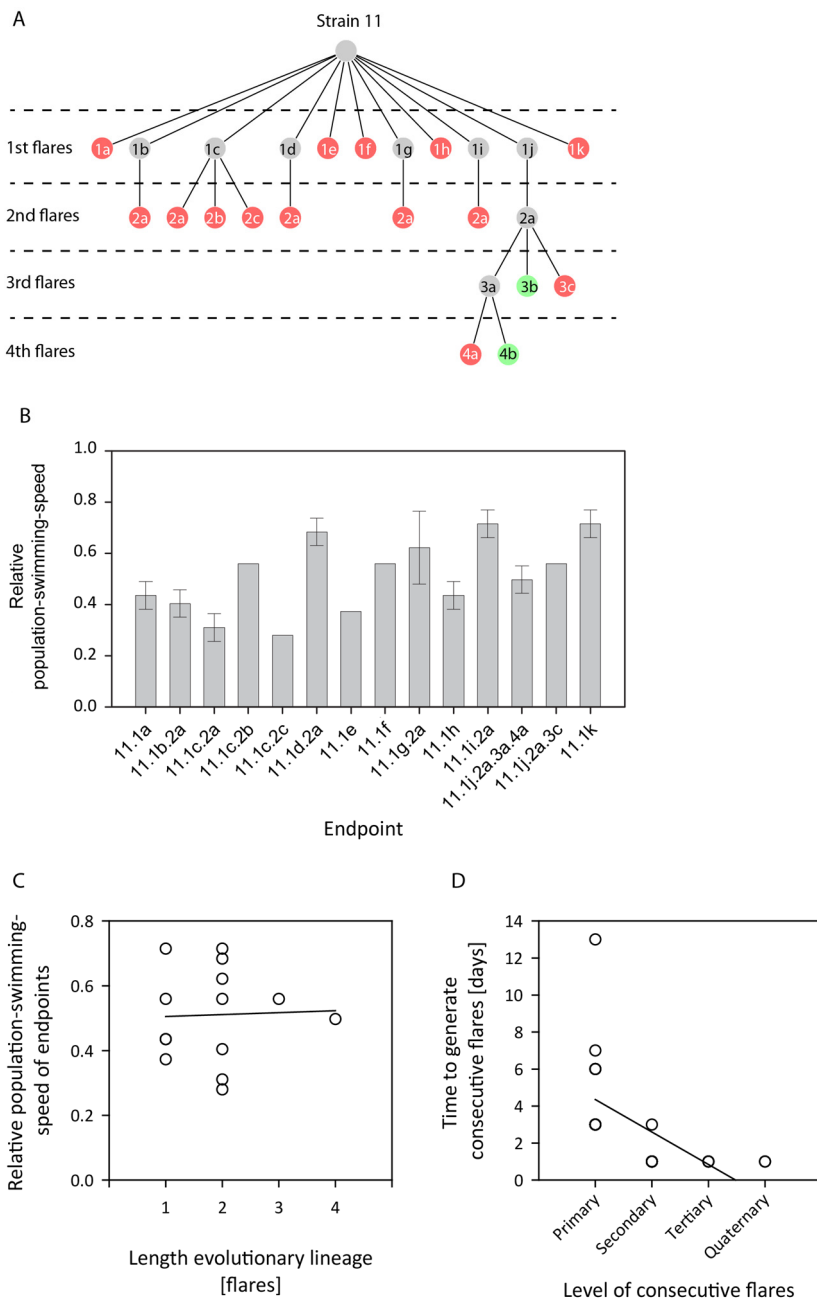


Figure 2.4. Evolution of improved motility in strain 11. **A.** Evolutionary lineages. Endpoints (genotypes isolated from flares that did not evolve improved motility) that were sequenced are indicated in red. Endpoints not sequenced are indicated in green. For nomenclature, see *Materials and Methods*. **B.** Population swimming speeds, relative to the reference strain harbouring the wild-type *E. coli* stators (strain 01). $n = 3$. Error bars indicate standard deviation. **C.** Correlation between the length of evolutionary lineages and the population swimming speeds of the corresponding endpoints relative to the reference strain. Linear correlation for illustrative purposes only. **D.** Correlation between the rank of flares and the time required for their generation. Linear correlation for illustrative purposes only. See text for statistics.

2.3.3.2 Strain 17

All 10 replicate populations of strain 17, which harboured stator genes from *Listeria monocytogenes* EGD-e, generated at least one primary flare between 2 and 3 days of selection after inoculation. Similar to strain 11, strain 17 also generated a diverse set of evolutionary lineages in terms of length and branching. However, as is clear from

Figure 2.5A, the dynamics were vastly different. Genotypes isolated from 12 of the 15 primary flares did not give rise to additional flares, while genotypes selected from the remaining 3 primary flares (17.1d, 17.1g and 17.1j) evolved improved motility and branched to great extents.

Statistical analysis demonstrated that the relative population swimming speeds were significantly different between endpoints (Figure 2.5B; Benjamini-Hochberg-corrected Kruskal-Wallis test; $n = 3$, $p < 0.001$), which indicated that different mutational trajectories to improved motility were taken. Unlike strain 11, strain 17 showed a significant, negative correlation between relative population swimming speeds of the endpoints and the length of evolutionary lineages (Figure 2.5C; Benjamini-Hochberg-corrected Spearman's correlation; $\rho = -0.669$; $p < 0.001$). This result indicated that, on average, the highest population swimming speeds were reached by mutational trajectories of few mutational steps, indicating that epistatic constraints were at play in the longer trajectories. Finally, a significant, negative correlation was found between the rank of a flare and the time required for its generation (Figure 2.5D; Spearman's correlation; $\rho = -0.815$; $p < 0.001$).

2.3.3.3 Strain 18

The last of the three cBFM strains that harboured compatible stators and evolved improved motility in the selection experiment is strain 18, which contained a sodium-driven stator complex from *Bacillus megaterium* DSM319 [61]. Between 2 and 6 days of selection, all 10 replicate populations had generated at least one primary flare, of which many contained genotypes that generated subsequent flares (Figure 2.6A). Only genotype 18.1m did not give rise to additional flares.

Strain 18 generated the largest number of flares compared with the other five strains that evolved (improved) motility during the selection experiment (see also section 2.3.4). In contrast to strain 17, the pattern of flares that evolved from strain 18 was more evenly distributed over the replicate lineages, as genotypes from many primary flares yielded secondary flares and genotypes from many secondary flares yielded tertiary flares (compare Figure 2.5A with Figure 2.6A). Between endpoints, significantly-different relative population swimming speeds were discovered (Figure 2.6B; Benjamini-Hochberg-corrected Kruskal-Wallis test; $n = 3$, $p < 0.001$). While strain 11 did not show a significant correlation between relative population swimming speed of endpoints and the length of evolutionary lineages, strain 17 showed a negative correlation. In contrast to both, we found a significant, positive correlation for strain 18 (Figure 2.6C; Benjamini-Hochberg-corrected Spearman's correlation; $\rho = 0.401$; $p = 0.013$). This indicated that, on average, higher population swimming speeds were reached after more mutational steps. Finally, a significant, negative correlation was found between the rank of consecutive flares and the time it took for them to evolve (Figure 2.6D; Spearman's correlation; $\rho = -0.573$; $p < 0.001$).

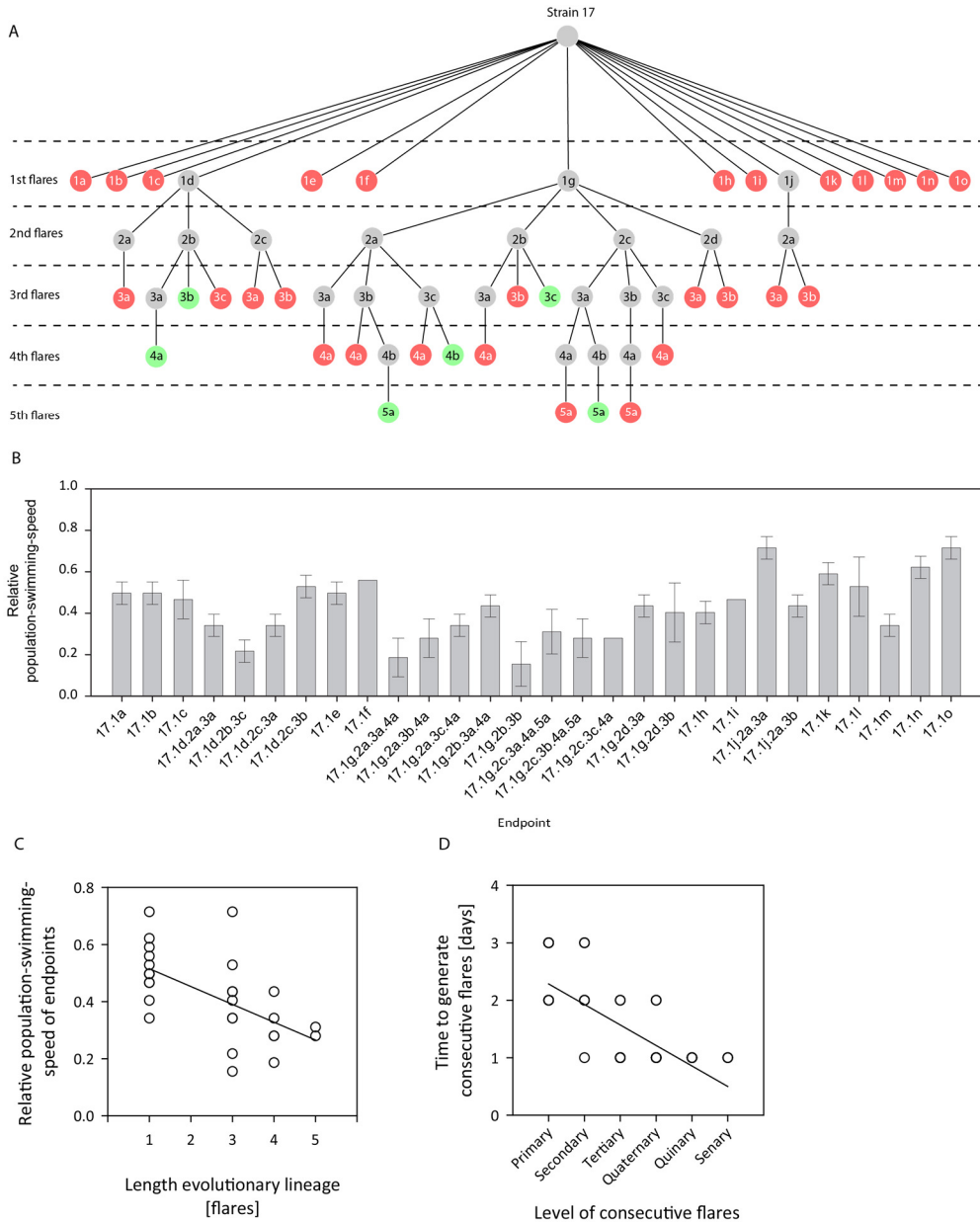


Figure 2.5 Evolution of improved motility in strain 17. A. Evolutionary lineages. Endpoints (genotypes isolated from flares that were not able to evolve improved motility) that were sequenced are indicated in red. Endpoints not sequenced are indicated in green. For nomenclature, see *Materials and Methods*. **B.** Population swimming speeds, relative to the reference strain harbouring the wild-type *E. coli* stators (strain O1). $n = 3$. Error bars indicate standard deviation. **C.** Correlation between the length of evolutionary lineages and the population swimming speeds of the corresponding endpoints relative to the reference strain. Linear correlation for illustrative purposes only. **D.** Correlation between the rank of flares and the time required for their generation. Linear correlation for illustrative purposes only. See text for statistics.

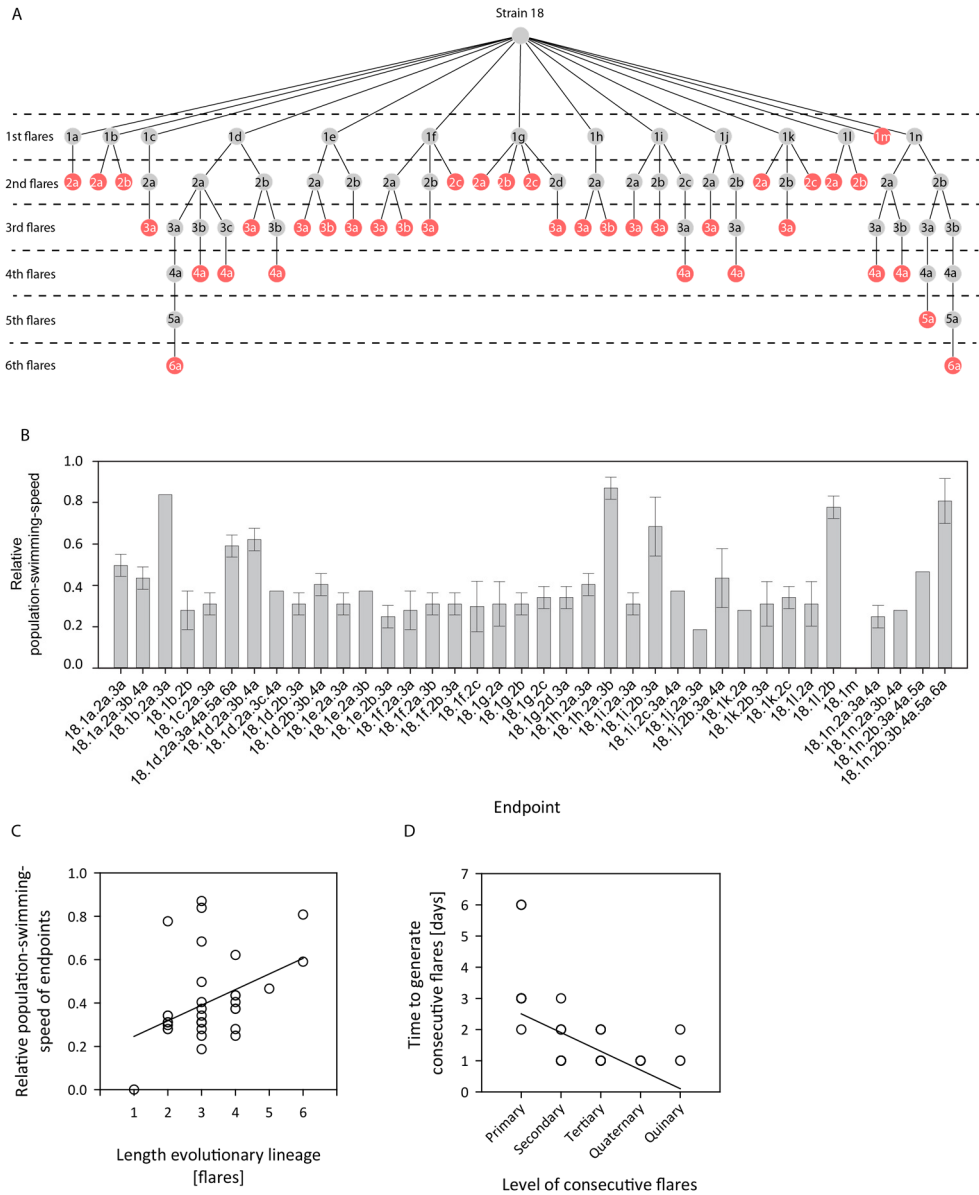


Figure 2.6 Evolution of improved motility in strain 18. **A.** Evolutionary lineages. Endpoints (genotypes isolated from flares that were not able to evolve improved motility) that were sequenced are indicated in red. For nomenclature, see *Materials and Methods*. **B.** Population swimming speeds, relative to the reference strain harbouring the wild-type *E. coli* stators (strain 01). $n = 3$. Error bars indicate standard deviation. Motility of 18.1m was too slow to be quantified with our method. **C.** Correlation between the length of evolutionary lineages and the population swimming speeds of the corresponding endpoints relative to the reference strain. Linear correlation for illustrative purposes only. **D.** Correlation between the rank of flares and the time required for their generation. Linear correlation for illustrative purposes only. See text for statistics.

2.3.4 Evolutionary integration of incompatible stators

In this section we discuss the strains of compatibility group C in more detail, which evolved compatibility of the foreign stators during the selection experiment followed by evolution of improved motility. These strains (08, 13 and 29) harboured stators that were components of lateral flagella in the donor species and the stators showed a percentage amino acid identity relative to WT stators ranging from 34% to 39%, in this respect being in between group A and B. The phylogenetic class of the gram-negative donor species was either γ -proteobacteria or α -proteobacteria.

2.3.4.1 Strain 08

Strain 08 harboured stator genes from the same donor species as from which the stator genes of strain 02 were derived. This donor species is *E. coli* O111:H – str. 11128 and it possesses two sets of stator genes. The genes introduced in strain 02 encoded stator proteins involved in peritrichous flagella, whereas those introduced in strain 08 encoded stator proteins involved in lateral flagella of the donor *E. coli* species (microbesonline.org).

During 35 days of selection, only 3 out of 10 replicate populations of strain 08 evolved compatibility of the stators. This yielded 5 primary flares (Figure 2.7A), of which two (8.1a and 8.1e) gave rise to genotypes that produced subsequent flares. The evolutionary lineage that started with genotype 8.1e branched and yielded five endpoints.

Statistical analysis demonstrated significant differences between the relative population swimming speeds of endpoints (Figure 2.7B; Benjamini-Hochberg-corrected Kruskal-Wallis test; $n = 3$, $p = 0.002$), which indicated that lineages followed different mutational trajectories. Similar to strain 17, a significant, negative correlation was found between the length of evolutionary lineages and relative population swimming speeds of endpoints (Figure 2.7C; Benjamini-Hochberg-corrected Spearman's correlation; $\rho = -0.719$; $p = 0.029$). The correlation between the rank of consecutive flares and the time required for their generation was negative (Figure 2.7D; Spearman's correlation; $\rho = -0.916$; $p < 0.001$).

2.3.4.2 Strain 13

Strain 13, which harboured stator genes from *Photobacterium profundum* SS9, required the most time to generate primary flares of the cBFM strains that evolved compatibility. Six replicate populations evolved compatibility between 13 and 27 days of selection, while four populations remained non-motile during 35 days of selection. As can be seen in Figure 2.8A, branching did not occur in the majority of replicate evolutionary lineages, although most of the genotypes isolated from primary flares did give rise to additional flares. Only one evolutionary lineage showed branching and this lineage yielded 50% of the endpoints.

Relative population swimming speeds differed significantly between endpoints (Figure 2.8B; Benjamini-Hochberg-corrected Kruskal-Wallis test; $n = 3$, $p = 0.002$), and, similar to strain 18, a significant, positive correlation between the length of evolutionary lineages and the relative population swimming speed of endpoints was observed (Figure 2.8C; Benjamini-Hochberg-corrected Spearman's correlation; $\rho = 0.651$; $p = 0.022$). The rank of consecutive flares correlated negatively

with the time to generate the consecutive flares (Figure 2.8D; Spearman's correlation; $\rho = -0.699$; $p < 0.001$).

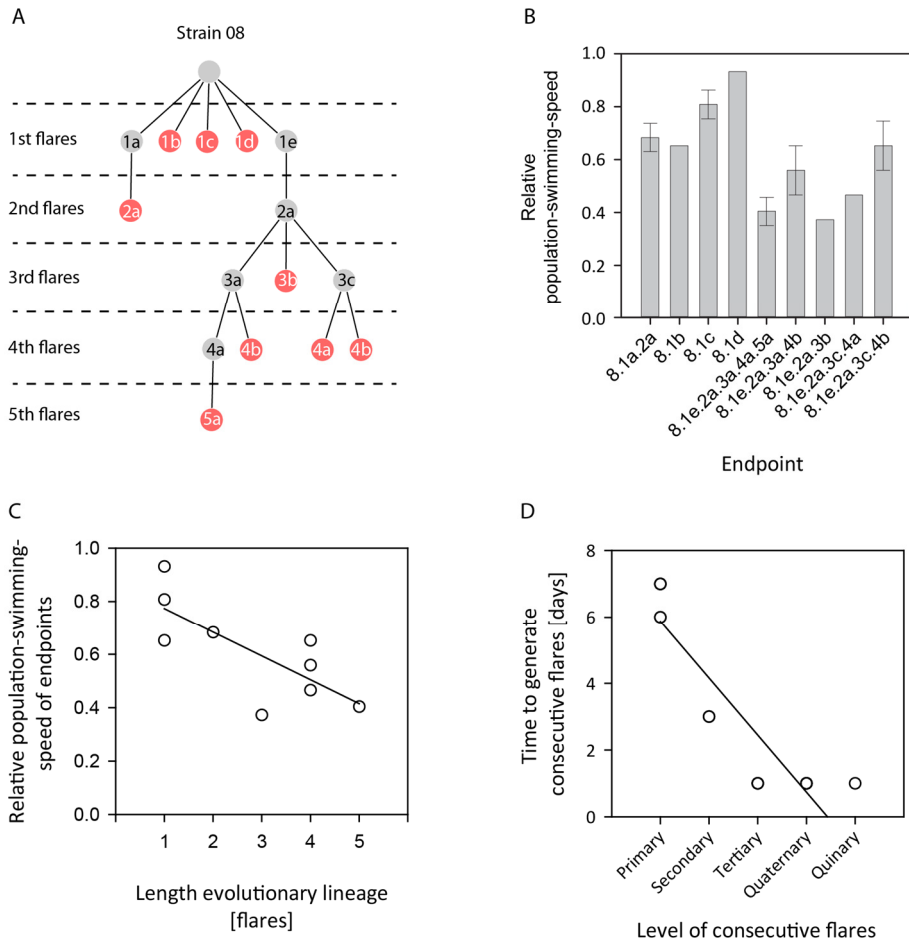


Figure 2.7. Evolution of compatibility and improved motility in strain 08. **A.** Evolutionary lineages. Endpoints (genotypes isolated from flares that were not able to evolve improved motility) that were sequenced are indicated in red. For nomenclature, see *Materials and Methods*. **B.** Population swimming speeds, relative to the reference strain harbouring the wild-type *E. coli* stators (strain 01). $n = 3$. Error bars indicate standard deviation. **C.** Correlation between the length of evolutionary lineages and the population swimming speeds of the corresponding endpoints relative to the reference strain. Linear correlation for illustrative purposes only. **D.** Correlation between the rank of flares and the time required for their generation. Linear correlation for illustrative purposes only. See text for statistics.

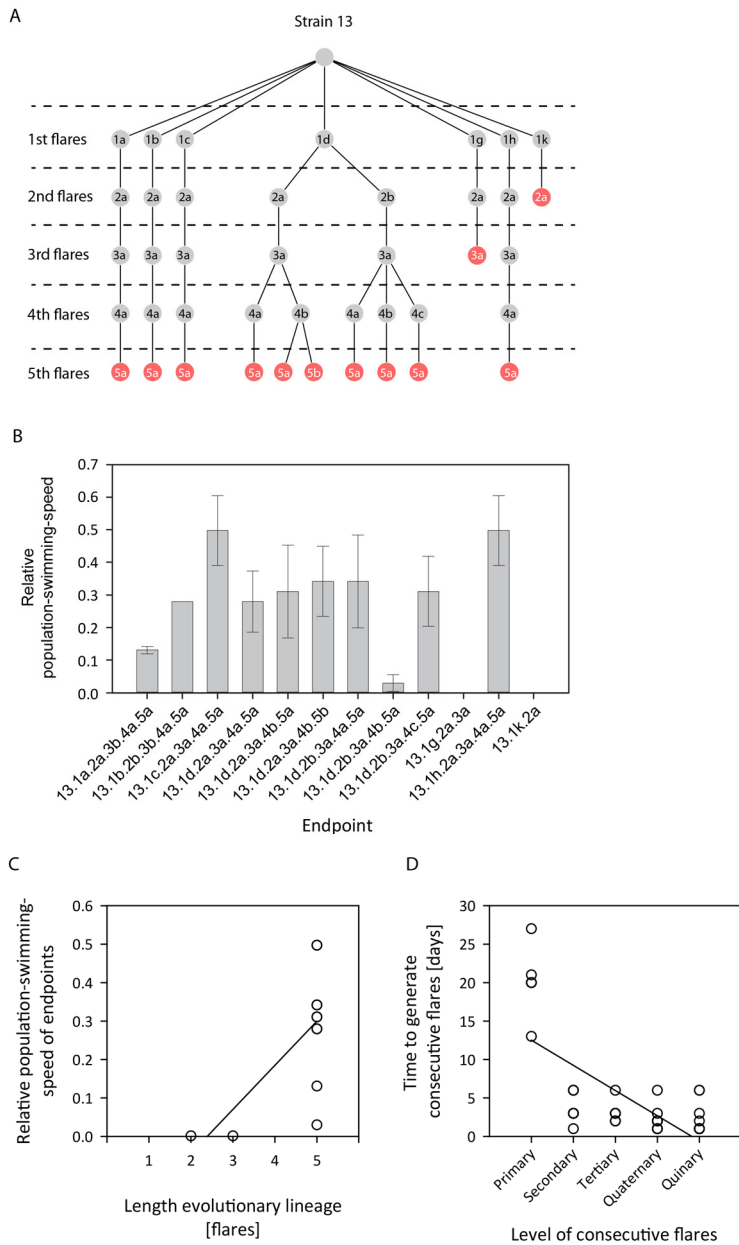


Figure 2.8 Evolution of compatibility and improved motility in strain 13. **A.** Evolutionary lineages. Endpoints (genotypes isolated from flares that were not able to evolve improved motility) that were sequenced are indicated in red. For nomenclature, see *Materials and Methods*. **B.** Population swimming speeds, relative to the reference strain harbouring the wild-type *E. coli* stators (strain 01). $n = 3$. Error bars indicate standard deviation. Motilities of 13.1g.2a.3a and 13.1k.2a were too slow to be quantified with our method. **C.** Correlation between the length of evolutionary lineages and the population swimming speeds of the corresponding endpoints relative to the reference strain. Linear correlation for illustrative purposes only. **D.** Correlation between the rank of flares and the time required for their generation. Linear correlation for illustrative purposes only. See text for statistics.

2.3.4.3 Strain 29

Strain 29 harboured stator genes of *Rhodospirillum centenum* SW, a donor species from the phylogenetic class α -proteobacteria. Within 13 days of selection, 9 out of 10 replicate populations evolved compatibility, which indicated more frequent evolution of compatibility compared with strains 08 and 13. Furthermore, strain 29 also yielded more endpoints (Figure 2.9A). The diversity among evolutionary lineages, both in terms of length and branching, indicated that different mutational trajectories to (improved) motility were taken. The observed statistically-significant differences between the relative population swimming speed of endpoints was consistent with this idea (Figure 2.9B; Benjamini-Hochberg-corrected Kruskal-Wallis test; $n = 3$, $p < 0.001$). Investigating the correlation between the length of evolutionary lineages and relative population swimming speed of endpoints, showed a significant, positive correlation (Figure 2.9C; Benjamini-Hochberg-corrected Spearman's correlation; $\rho = 0.541$; $p = 0.004$), which indicated that longer evolutionary lineages resulted, on average, in endpoint genotypes that had higher relative population swimming speeds. Similar to all strains discussed above, a significant, negative correlation was found between the rank of flares and the time to generate those flares (Figure 2.9D; Spearman's correlation; $\rho = -0.542$; $p < 0.001$).

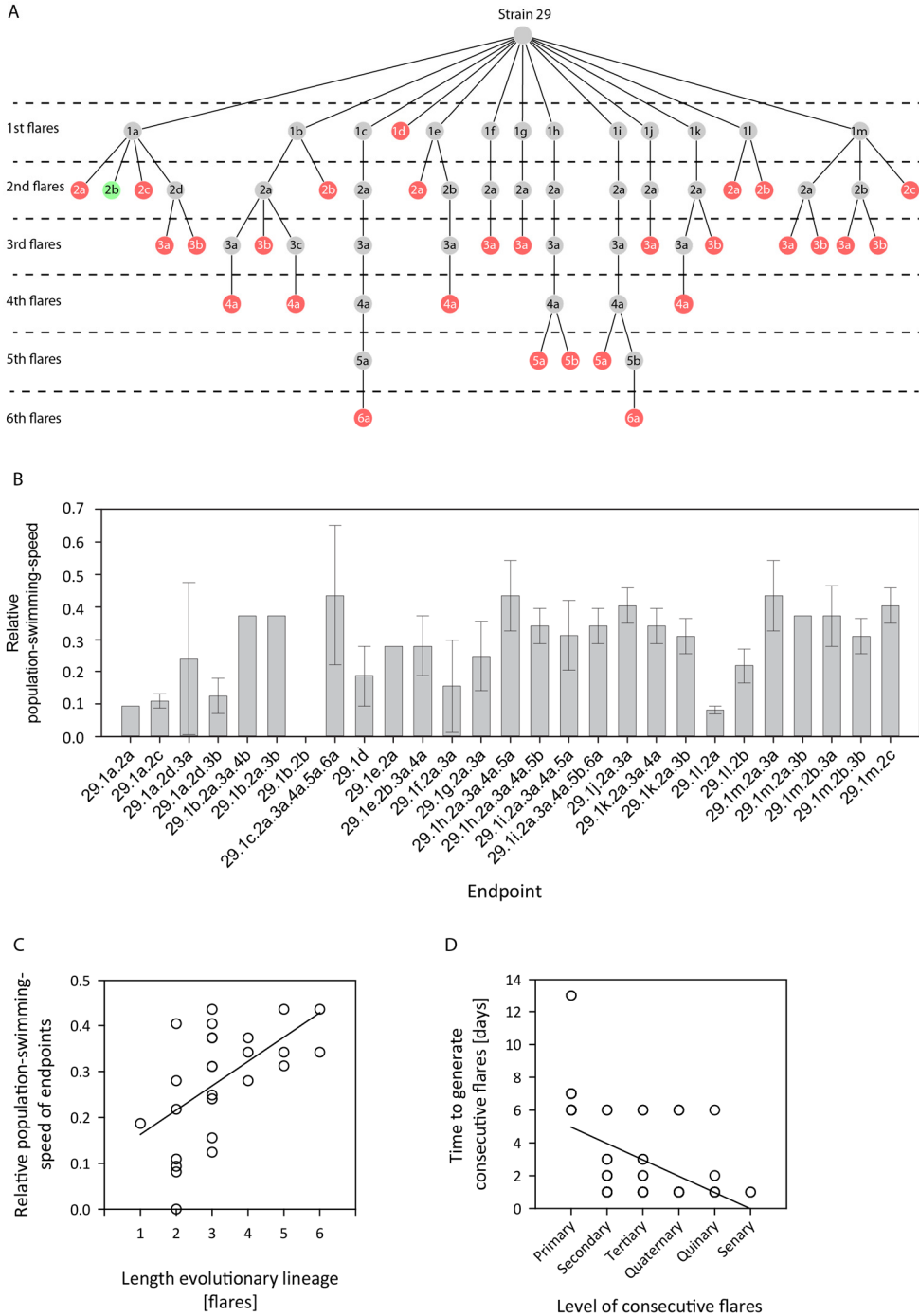


Figure 2.9 Evolution of compatibility and improved motility in strain 29. A. Evolutionary lineages. Endpoints (genotypes isolated from flares that were not able to evolve improved motility) that were sequenced are indicated in red. Endpoints not sequenced are indicated in green. For nomenclature, see *Materials and Methods*. **B.** Population swimming speeds, relative to the reference strain harbouring the wild-type *E. coli* stators (strain O1). $n = 3$. Error bars indicate standard

deviation. Motility of 29.1b.2b was too slow to be quantified with our method. **C.** Correlation between the length of evolutionary lineages and the population swimming speeds of the corresponding endpoints relative to the reference strain. Linear correlation for illustrative purposes only. **D.** Correlation between the rank of flares and the time required for their generation. Linear correlation for illustrative purposes only. See text for statistics.

2.3.4.4 Comparison of the integration dynamics of foreign stators

In the previous sections, we investigated for individual strains the evolution of (improved) motility during our selection experiment. In order to examine the differences between strains, this section compared the key characteristics of the evolutionary lineages of all strains (Table 2.2).

As indicated in section 2.3.1, we divided the strains that evolved (improved) motility into two groups. Compatibility group B contained strains that harboured stators that were compatible immediately, whereas compatibility group C contained strains with stators that were incompatible at the start. All 10 replicate populations of the strains that belonged to compatibility group B evolved improved motility, in contrast to the replicate populations of strains that belonged to group C. Furthermore, the strains of group C required more time to generate primary flares. These results demonstrated that different foreign stators engendered different initial constraints on the mutational trajectories that were available to increase motility.

All strains that evolutionarily integrated foreign stators, irrespective of the compatibility group, showed a significant, negative correlation between the rank of a flare and the time required for its emergence. Thus, overall, as the foreign stators became better integrated into the *E. coli* cell, mutant genotypes with a higher fitness fixed at a higher rate. We suggested that this was due to either a higher availability of mutations that could increase population swimming speed or due to a larger fitness effect of mutations.

The short average length of evolutionary lineages of strains 08, 11 and 17, in comparison to those of 13, 18, and 29, showed that their starting genotypes were differently constrained, as the former could reach genotypes with a high fitness with fewer mutational steps than the latter. This indicated that the latter strains lacked access to *big-benefit* mutations, which facilitated a large increase in relative population swimming speed, early in their evolutionary trajectory. Here, we consider a 'large increase' to be an increase in relative population swimming speed of >50% of the difference between the highest speed obtained by an endpoint in that strain and the initial speed. The notion that strains 13, 18 and 29 lacked access to big-benefit mutations early on is further supported by the observation that, in these lineages, there was a positive correlation between the length of an evolutionary lineage and the relative population swimming speed of the endpoint genotype. This showed that these strains required more mutations to reach a high fitness. The negative correlation found for strains 08 and 17 indicated that longer evolutionary lineages did not reach the highest fitness among replicate lineages. This suggested that some of the longer trajectories of these two strains (the trajectories that did not reach high fitness) did so owing to constraints engendered by epistatic interactions with the earlier mutations. These epistatic interactions rendered the big-benefit mutations that fixed in the shorter lineages either less effective (magnitude epistasis) or not accessible (sign-epistasis) [64]. No significant correlation between trajectory length and end-point fitness was discovered for strain 11.

2.3.5 Stators beyond the compatibility horizon

Compatibility group D (Table 2.1) contained strains that harboured incompatible stators and that did not evolve compatibility during 35 days of selection. Because the amino acid identity of the stators compared with WT *E. coli* stators ranged from 24.5% to 42%, the strains of compatibility group D overlapped with the strains of both groups B and C in this respect. The donor strains of the stator genes included Firmicutes, α -, γ - and δ -proteobacteria, and the positioning of flagella in these strains was either polar, lateral or peritrichous. Qualitative video-tracking of the strains from group D confirmed the non-motile phenotype (data not shown).

Table 2.2. Key characteristics of all strains that evolved (improved) motility during the selection experiment.

Compatibility group	Strain	Number of replicate populations that evolved (out of 10)	Average time to 1st flare [days]	Correlation level consecutive flares and time to generate consecutive flares	Average length evolutionary lineage per primary flare [flares]	Correlation length evolutionary lineage and relative population-swimming-speed of endpoints
B	11	10	5.1	negative	1.68	no correlation
	17	10	2.5	negative	1.47	negative
	18	10	3.6	negative	2.87	positive
C	08	3	6.6	negative	1.80	negative
	13	6	20.1	negative	4.29	positive
	29	9	7.4	negative	3.34	positive

2.4 Discussion

In spite of the central role of compositional evolution in explaining the origin, innovation and adaptation of protein complexes, empirical insight into its creative potential and phenotypic dynamics is lacking. In this chapter, we described an evolutionary experiment that exploited cells with engineered BFMs, poised to integrate a foreign component, to address this important issue. The experiment successfully captured evolutionary integration events and indicated the presence of a compatibility horizon, i.e. a connectional zone in the spectrum of foreign components that separates incompatibilities that can be resolved by step-wise evolution from those that cannot. As such, it provided the starting material that will be used throughout the following three chapters of this thesis, to shed light on this modular mode of protein-complex evolution. This chapter examined the phenotypic dynamics and inferred from these a number of basic characteristics about the nature of compatibility of homologous BFM components and the fitness landscape that governs their integration. The results opened a unique window on this process and showed that immediate compatibility was not limited to a narrow phylogenetic range. Rather, compatibility correlated positively with the degree of structural similarity between the BFM of the donor species of the foreign component and the recipient BFM. Introduction of foreign components generated significant heritable, phenotypic variation and included cBFMs that endowed the host cells with a superior, fitness-related phenotype in our selective environment, a requirement for innovation and adaptation by modular adaptive evolution. Within-strain diversity in evolutionary lineages that underwent evolutionary integration of the foreign components revealed the existence of alternative, mutationally-accessible trajectories along which integration could occur and thus revealed differences in the structure of the local fitness landscapes. Furthermore, different correlations between the length of evolutionary lineages and the relative population swimming speed of endpoints indicated differences in local fitness landscape between different engineered strains. Our findings strongly indicate a role for epistatic interactions as a factor that constrains the degree of evolutionary component integration that is achievable in the short term.

2.4.1 Predictors of compatibility and evolutionary integration

The percentage amino acid identity relative to *E. coli* WT stators of the compatible, foreign stators (compatibility group A and B) varied between 99.7% to 23.8%. Although these percentages differed significantly from those of incompatible stators (Table 2.1; compatibility group C and D), significant differences in amino acid percentage identity were also discovered between compatibility group A and B. Therefore, amino acid identity was not a reliable predictor of compatibility. The most accurate predictor of immediate compatibility that we identified was the flagellation type of the donor species of the foreign stator. If these cells were peritrichously flagellated, as is the case for the *E. coli* host cells, then, at least for eight out of nine cases we examined, the foreign stator was compatible with the *E. coli* system. Because BFM structures differ between different types of flagellation [68], structural differences between stator complexes might underlie the initial compatibility. For instance, interaction between the stator complex and the rotor is necessary for flagellar rotation [24,26]. Misalignment between rotor and stator due to differential interactions of the stator complex with the peptidoglycan layer or a poor fit of the stator complex with the BFM, might thus result in incompatibility.

Between different cBFM strains with compatible stators, different population phenotypes were observed. Strains of compatibility group A, which did not show evolutionary integration of foreign stators, showed motility similar to a strain harbouring WT *E. coli* stators, whereas strains of compatibility group B, which did show evolutionary integration, showed ATM (Figure 2.3). Key characteristics of the (donor species of the) foreign stators that differed between these groups, and might therefore be important in determining the phenotype of motility, were similarity in amino acid sequence, the phylogenetic class of the donor species and gram-staining of the donor species. Because sequence similarity was revealed as a poor predictor of initial compatibility, it will not be discussed further. However, both the phylogenetic class and gram-staining of the donor species might provide additional insight into the differences in motility. As different classes of bacteria are known to have different BFM structures [68], a different fit of the foreign stators with the *E. coli* BFM might be expected between stators from different phylogenetic classes. Differences in gram-staining indicate differences in the composition of the cell wall of bacteria. As the cell wall has an interaction with the stator complex (peptidoglycan-MotB interaction [69,70]), a different composition of the cell wall of the donor species might lead to differently-adapted stator complexes and thus a sub-optimal fit with the *E. coli* BFM.

No differences between the key characteristics of the (donor species of the) foreign stators were observed between strains of compatibility group C and D, i.e. strains that harboured incompatible stators that were evolutionarily integrated and ones that were beyond the compatibility horizon, respectively. Because many donor species of the stators in group D have polar flagella instead of lateral flagella that were found in the donor species of group C, differences in the fit between stators and the *E. coli* BFM might explain the observed differences in the capacity to be evolutionarily integrated.

In our experimental system, multiple possibilities might underlie initial compatibility between foreign stators and the *E. coli* system. One of these possibilities was mentioned above and is represented by structural differences between the foreign stators and the *E. coli* WT stator. For instance, as a result of these structural differences, foreign stators might not have an optimal fit with the *E. coli* BFM. Other possibilities that might have influenced initial compatibility of the foreign stators as well as the capacity of evolution to achieve (improved) functionality, are factors that influenced transcription and translation, such as differences in the Shine-Dalgarno (SD) sequence of MotB, differences in GC-percentage of stator genes and codon usage. As a consequence of our engineering strategy was the SD sequence of MotA homologues identical for all cBFM strains and consisted of the *E. coli* MotA SD sequence. However, the SD sequence of MotB homologues was different for cBFM strains and consisted of the MotB SD sequence of the respective donor species. If this SD sequence was incompatible with the translational machinery of *E. coli*, it might have caused MotB genes to be not or sub-optimally translated (e.g. [71]), which might have led to the absence of MotB proteins or presence at very low levels, respectively. Subsequently, this could have had an effect on BFM-mediated motility. Differences in GC-percentage (although not significantly different between compatibility groups) and differences in codon usage [72,73] might have had similar effects on the levels of both stator proteins. Thus, in addition to structural differences, transcriptional and translational issues might also have determined compatibility of the foreign stators. Further mechanisms that may have been of influence on compatibility and evolutionary integration are

efficiency of membrane insertion of the foreign stators, interaction of the BFM with the chemosensory pathway or other cell systems and the ability of the stators to use the PMF of *E. coli* for motility. Especially the latter possibility might underlie the high percentage of non-motile cells in both strains 11 and 18. In the donor species of strain 11 (*Bacillus pseudofirmus* OF4), protons cannot be used for motility and are hypothesized to bind competitively to the stator complex [59–61], interfering with sodium-ion binding which drives motility. The same mechanism may hold for the stators of strain 18. As the semi-solid agar medium contains 85mM NaCl at neutral pH (*Materials and Methods*), this concentration might just be enough to cause motility occasionally, leveraging at the boundary of efficient competitive inhibition by protons [59]. Experiments using the sodium-channel blocker EIPA, titrating NaCl-concentrations or changing the pH of the medium could shed light on the ability of both stators to use protons to facilitate flagellar rotation.

Taken together, our results revealed that stators covering a broad phylogenetic range were compatible with *E. coli* or could be compatibilized by evolution to function in a foreign system. The functional and structural factors that constrained direct compatibility and evolutionary integration represent evolutionary limits, that likely shape the flow of genes between bacterial species through horizontal gene transfer [38,74]. The findings illustrate for this particular model, a large potential for evolutionary adaptation and innovation by compositional evolution, which is consistent with the role of this mode of evolution in the evolution of complex molecular machines and protein complexes in general.

2.4.2 Phenotypic diversity caused by stator exchange

One requirement for compositional evolution is that the foreign components are either directly compatible or that evolution can generate compatibility. Especially the former mechanism might be very promising, as the phenotypic diversity generated can be acted upon immediately by natural selection. When the environment is constant, organisms with foreign components must be fitter than the ancestor to be selected. In other words, the diversity generated must be adaptive in the current environment. However, when the environment changes, which is often the case in natural environments, any phenotypic diversity generated may turn out to be adaptive. Our results showed that 8 of the 22 foreign stators were compatible immediately with the *E. coli* cell and BFM (Table 2.1; compatibility group A and B). Of these foreign stators, three resulted in strains with a similar or higher population motility compared with a strain harbouring WT stators, whereas the other strains showed lower motility (Figure 2.1 and Figure 2.3). At the single-cell level, phenotypic variation was observed for the swimming speed (Figure 2.2). Together, these results indicated that adding foreign components can result immediately in heritable phenotypic variation at multiple levels of biological organization, which may be selected in an environment where it is favoured.

For two strains that harboured stators that were compatible directly, we measured three single-cell phenotypes (single-cell swimming speed, percentage of non-motile cells in the population, cell size) that might underlie the differences observed at the population level described above. As such, they provide deeper insight into the mechanisms behind the phenotypic effects of introducing foreign stators (Figure 2.2). One of these strains (strain 24) showed a population motility that was higher than that for a strain with wild-type stators, whereas the other strain (strain 26) showed a population motility that was lower than the strain with wild-type stators. To be able to understand

how the single-cell phenotypes might affect population motility, it is essential to dissect the driving force behind population motility. The driving force behind population motility in our selective environment (semi-solid agar) is chemotaxis. Chemotaxis is the ability of cells to sense and move away from a repellent gradient or to follow an attractant gradient (e.g. nutrients) [30]. The generation of such nutrient gradients, which have a sigmoidal shape [75], and their steepness are dependent on the rate at which the nutrient is consumed by a population of cells. This, in turn, is dependent on several physiological factors, which include consumption rate by individual cells, growth rate of individual cells and population size. Decreases in the rate of these cellular parameters will result in the slower generation of the gradient as well as a less steep gradient, which on its turn will negatively affect the rate of chemotaxis [53]. Below we will discuss three single-cell phenotypes that might also affect chemotaxis and hence population motility.

A different percentage of non-motile cells in the population might affect population motility by influencing the size of the population that performs chemotaxis. For instance, a higher percentage of non-motile cells leads to less cells (a smaller population) that will follow the self-created nutrient gradient. In turn, this will result in a nutrient gradient that will be formed more slowly (less cells consuming nutrient), which will affect the rate of chemotaxis negatively. Similarly, a lower percentage of non-motile cells might lead to a higher population motility. Because strain 26 showed a higher percentage of non-motile cells than the strain harbouring wild-type stators, the lower population motility of this strain might thus be explained.

Differences in single-cell swimming speed might also affect population motility. As cells move outward from the inoculation site and continue to consume the nutrient, the nutrient gradient also moves outward. If cells are not able to follow the moving nutrient-gradient efficiently due to lower single-cell swimming speeds, this might result in cells not sensing the steepest part of the gradient [75], which will result in slower chemotaxis. Similar to what was discussed before for the percentage of non-motile cells in the population, a lower population motility of strain 26 may be explained by its lower single-cell swimming speed.

As the pore size of semi-solid agar is in the range of the width of *E. coli* cells [76–79], cell size might influence the ability of cells to migrate through these pores and might thus affect population motility. However, this factor could be excluded for strain 24 and 26, as cell size did not change upon component swapping for the two strains tested.

Overall, this analysis showed that foreign components could generate heritable phenotypic diversity at several levels, which included a population-level chemotactic motility at a rate higher than that of a strain with the wild-type stators. It furthermore demonstrated that, for instance in cases such as strain 24, phenotypes other than the single-cell swimming speed or percentage of non-motile cells in the population may underlie differences observed at the population level. For instance, properties of the foreign component may change the interaction between the BFM and the chemotaxis signalling pathway, thereby changing tumbling frequency [33,80,81] and chemotactic efficiency.

2.4.3 Phenotypic dynamics of integration and the fitness landscape

We aimed to address four questions regarding the phenotypic dynamics associated with evolutionary integration of a foreign component into a protein complex. Here we will discuss the data in the light of these questions.

2.4.3.1 Multi-step trajectories

Our first question was whether the complexity of the interactions between the components of the BFM would allow for foreign component integration along multi-step, mutational trajectories. This question was fuelled by the following two arguments, that make evolution of protein complexes by compositional evolution seem unlikely. The first argument is that protein complexes often contain protein parts that are essential to it. Without these parts, the protein complexes do not function. Therefore, some argue that protein complexes are “irreducibly complex” [5,82] and therefore cannot have evolved. The second argument is the apparent presence of many strong and specific interactions between protein-complex components (epistasis). This might result in the necessity of multiple beneficial mutations simultaneously in several protein parts to improve function, which lowers the chance of improving function by mutations [17,19,20,83,84]. However, our results prove otherwise. Although we cannot pinpoint the location of mutations that caused evolutionary integration at this point, we can conclude that multi-step trajectories on the underlying fitness landscape were possible. In fact, all six strains that showed evolutionary integration of the foreign stators, of which three contained incompatible foreign stators, contained at least two trajectories that were composed of multiple steps. The results thus demonstrated that, despite the presence of epistasis between BFM components, evolution by a series of point mutations was possible.

2.4.3.2 Alternative trajectories

The second question that we raised was whether alternative evolutionary trajectories existed along which evolutionary integration could occur in a strain. We addressed this question by using replicate selection lines for each engineered strain and by allowing branching of the resulting lineages. All cBFM strains that integrated foreign stators by evolution, showed a diversity in both length and branching pattern of evolutionary trajectories, as well as diversity in relative population swimming speed of endpoints. These results indicated that multiple mutational routes on the fitness landscape were selectively available. So, despite the presence of epistatic interactions between protein parts of the BFM, we did not only find that foreign components could be integrated via multiple steps, but also that different routes were possible. In other words, several multiple-step solutions existed to functionally integrate a foreign component, which was suggested to be very unlikely by others [5,17,20,82–84]. Although different mutations by definition have different locations on the underlying fitness landscape, we cannot draw conclusions about the ruggedness of the fitness landscape [64]. On the one hand, it might be possible for all trajectories to be located on one fitness peak, which converges to a single top or plateau. Alternatively, it might be possible that the fitness landscape is composed of different local optima that are separated by fitness valleys. As examining the entire fitness landscape is beyond the scope of this thesis, we will not go into more detail on this topic.

2.4.3.3 Epistasis

The third question examined the influence of epistasis on the genetic trajectories available for evolutionary integration of a foreign component. An inevitable limitation of complex systems, in which function emerges from the interactions between components, is that the effect of one change may depend on the state of the entire system—that is, evolution may be shaped by epistasis. Furthermore, as mentioned above, some argue that epistasis might effectively prevent the evolution of protein complexes by point mutations as a result of the necessity of multiple mutations simultaneously to improve function [17,19,20,83,84]. While we have seen in our experimental system that evolutionary integration of foreign components was not prevented by epistasis, we did not examine its influence so far.

One way to examine epistasis in the phenotypic data set described in this chapter is to identify, for strains that give rise to alternative adaptive trajectories that differ in length and endpoint fitness, the correlation between the number of steps and the endpoint fitness. A negative correlation indicates that, on average, high fitness can be reached with one step, but not with many steps. We suggest that high fitness in these cases cannot be reached with many steps due to either sign epistasis or magnitude epistasis [64]. Our results revealed a negative correlation between trajectory length and endpoint fitness for two cBFM strains. This indicated that, although BFM complexity allowed for multiple step-wise trajectories along which a foreign component could be integrated, compositional evolution was constrained by epistasis.

Another way to examine the influence of epistasis, although more indirectly, was to investigate the repeatability of evolutionary integration of foreign stators in compatibility group C (Table 2.1). The three strains in compatibility group C were all non-motile initially, which indicated that the stators were incompatible, and did not show evolutionary integration of the stators in all replicate selection lines. When non-motile, cells could evolve a higher ability to spread out in semi-solid agar by, for instance, increasing growth rate [48]. Given our selection regime (*Materials and Methods*), the increased ability to spread out in semi-solid agar would have increased the chance of this genotype being selected, either as a flare or as a large part of the transferred population. These mutations, although increasing fitness, might have negatively influenced the effects of mutations that caused compatibility otherwise (sign epistasis), or might have reduced the fitness effects of those mutations (magnitude epistasis). Another explanation for the fact that not all replicates evolved compatibility might be that several mutations were necessary for compatibilization. This would have lowered the chance of realization and is the mechanism that was proposed before to have prevented the evolution of protein complexes by point mutations [17,20,83,84].

Taken together, our results suggest that, depending on the identity of the foreign component, epistasis shapes its evolutionary integration. In some cases it might prevent evolution of protein complexes by compositional evolution, while in other cases it directs the mutational route that is taken.

2.4.3.4 Between strain comparison

The fourth question we asked was whether differences or similarities in the dynamics of evolutionary integration between cBFM strains could be discovered, besides epistasis mentioned in the previous paragraph. As different strains harboured different foreign stators, the strains

represented different starting points in genotype space and as such can provide information on the fitness landscape on a bigger scale.

2.4.3.4.1 Correlation flare rank and time to generate flare

We found a negative correlation for all cBFM strains between the rank of consecutive flares (primary, secondary, etc.) and the time to generate those flares. Especially the time required for the first flare to be generated was higher than for subsequent flares. This characteristic thus seemed unaffected by the position on the fitness landscape and indicated that with improved integration of foreign stators (i.e. improved population motility in semi-solid agar), mutations that increased population motility further fixed at a higher rate.

One possibility for this observation is that with improved integration of the foreign stators, more mutations became available to improve motility further. A possible mechanism for the increased availability of mutations is that the first mutations, which resulted in integration, especially improved the integration of the foreign stator proteins into the *E. coli* BFM or improved expression of the foreign stator genes (few loci affected). As the foreign stators were not adapted to *E. coli* before these mutations, there might have been some dominant conflicts with *E. coli* systems that needed to be resolved first. Here, dominant conflicts are defined as conflicts that need to be resolved before any other improvement can occur. Therefore, limited possibilities for improving population motility might have been possible. For instance, it is useless (and evolutionary unlikely) to improve the efficiency of chemotaxis, if the BFM is not providing (sufficient) cellular motility. Next, mutations that further improved population motility could have caused improved integration of the foreign stators into the BFM by affecting more loci than the initial mutations or could have affected less-specific loci, such as genes involved in the chemotaxis signalling pathway. After the first mutations, the foreign stators were more adapted to *E. coli*, which might have resulted in the absence of dominant conflicts with *E. coli* systems and more fruitful interactions. This, in turn, might have resulted in more possibilities to improve motility further. For instance, when the BFM functioned after the first mutation and resulted in cellular and population motility, further improvement of population motility might have occurred via improving BFM functioning (e.g. higher cellular motility — see section 2.3.2) or via improving chemotactic efficiency. Finally, mutations that resulted in even higher population motility might have affected many BFM-related loci and non-BFM related loci (e.g. loci that influence nutrient metabolism). After the second round of mutations that caused integration, it might have been possible that more fruitful interactions with *E. coli* systems had arisen, which could have led to more opportunities for improving population motility.

Another possibility for the increased fixation rate of mutations that increased population motility further, which is not mutually exclusive with an increased availability of mutations, is an increased fitness effect of mutations. Previously [66,67], it was demonstrated that the time to reach fixation was inversely proportional to the relative fitness advantage as a result of the mutation. In other words, the larger the fitness effect of a mutation, the shorter the time to fixate. However, other studies have indicated that the increments in fitness may become smaller with improved adaptation [85–87]. In our experiment, this would have resulted in a longer time to generate flares with improved integration, which is the opposite of what we found. We thus suggest a more prominent role for an increase in mutations that become available to improve population motility further.

Furthermore, an increase in the number of loci with a positive effect on fitness when mutated, such as suggested here, might be a general feature of compositional evolution. Initial mutations might be restricted to the foreign components and their interaction partners to fine-tune interactions and expression, whereas subsequent mutations can increase fitness by many more routes.

2.4.3.4.2 Evolutionary lineage length and big-benefit mutations

While different cBFM strains showed similarities in the fixation rate of mutations that improved motility, they differed in the average lineage length of their adaptive trajectories. This suggested that some cBFM strains (13, 18 and 29) required more mutational steps to reach higher motility than other strains (08, 11 and 17). The former strains especially lacked one-step evolutionary lineages to high motility and thus lacked the availability of big-benefit mutations early in evolution. We defined big-benefit mutations as mutations that led to an increase in relative population swimming speed of >50% of the difference between the highest speed obtained by an endpoint in that strain and the initial speed. A positive correlation between the length of evolutionary lineages and the relative population swimming speed of endpoints supported the idea of a lack of big-benefit mutations early in evolution for cBFM strains 13, 18 and 29. These results suggested that evolution was constrained by the position of strains in the fitness landscape, which determined the steepness of the slopes in the fitness landscape.

A possibility that might underlie these differences in lineage length and availability of big-benefit mutations early in evolution is that different foreign stators interacted differently with *E. coli* systems. In other words, different foreign stators had diverse incompatibilities that needed diverse solutions. Depending on the number of dominant conflicts and nature of fruitful interactions that occurred immediately upon replacing the *E. coli* WT stators and after the first mutations that led to functional integration, the foreign stators might have been integrated with few steps (e.g. no or very few dominant conflicts and many fruitful interactions) or many steps (e.g. quite some dominant conflicts and few fruitful interactions). If this were to be true, our results would indicate that, for this example of a protein complex, different types of incompatibilities could be resolved.

To summarize, a comparison between cBFM strains showed that the position in the fitness landscape might determine specific dynamics of evolutionary integration of foreign components. However, general characteristics of evolution that were independent of the position in the fitness landscape, namely the increase in fixation rate of beneficial mutations with improved integration of foreign components, were also found.

2.4.4 ATM

Some of the cBFM strains showed a specific type of population motility in semi-solid agar that was different from the chemotactic motility that was displayed by the reference strain harbouring WT *E. coli* stators. This type of motility was called ATM and was characterised by dots radiating from the point of inoculation (after 24 h; Figure 2.3). The possibility of ATM showed that compositional evolution could lead to intermediate motility levels, in between the absence of motility and the presence of normal chemotactic motility. As such, it represented an immediately selectable starting point for adaptive evolution by point mutations and thus facilitated compositional evolution.

2.4.4.1 Mechanisms underlying ATM

The ATM phenotype was previously reported to be caused by a low number of motile cells in the population that give rise to predominantly non-motile daughter cells [25,57,58]. When a cell swims away from the point of inoculation and yields non-motile daughter cells, which continue to grow and divide, a spherical bacterial colony is formed in the agar away from the point of inoculation. Because ATM is characterised by *satellite micro-colonies* [25,57,58] that surround the site of inoculation, a low number of motile cells that yield predominantly non-motile daughter cells might explain the ATM phenotype. But what mechanism might underlie the phenomenon of non-motile cells in a population, as also a population of cells with WT stators shows a (low) percentage of non-motile cells (Figure 2.2B)? A possible explanation is that large and long-lasting fluctuations in the number of stators per cell and stators docked in the BFM result in cell-to-cell variation with non-motile cells as a result. When stators are well integrated into the cell (e.g. transcription, translation and cellular interactions), resulting in high single-cell motility, the effects of fluctuations might be minimal. This could result in few non-motile cells. However, when stators are less well integrated into the cell, resulting in lower single-cell motility, the effects of fluctuations might be larger. This could result in a higher percentage of non-motile cells. Microscopic observations of strains 11 and 18 support the notion that the mechanism of few motile cells (with less-well integrated stators and thus lower single-cell motility) might underlie ATM in these strains. However, populations of strain 17 contained many motile cells (similar to WT level) that were faster than cells of strains 11 and 18, but slower than WT cells (observation). Thus, while a high number of non-motile cells might underlie ATM in strains 11 and 18, different factors may cause the ATM phenotype of strain 17. Factors that could explain this discrepancy include the absence of tumbling, a low tumbling frequency or low torque generated by the BFM. Note that these factors may also play a role in strains 11 and 18 and are not mutually exclusive with a high number of non-motile cells.

The tumbling frequency of the cells is not only important for chemotaxis, but also to escape from dead-end pores in semi-solid agar. Dead-end pores are formed in semi-solid agar during solidifying, when a fine network of agarose polymers [76] with pore sizes in the order of the cell width of *E. coli* [76–79] is established. Because tumbling allows a cell to re-orient itself (e.g. [33,80]) and is known to affect the movement of cells through agar [81], insufficient tumbling or the absence of it might lead to cells that cannot escape after being trapped in a dead-end pore. When cells continue to grow and divide when trapped, a spherical bacterial colony is formed, similar to what was described for non-motile cells. A possible mechanism behind the absence of tumbling or a low tumbling frequency is that a cBFM has no or a suboptimal interaction with the chemotaxis signalling pathway, for instance due to structural incompatibilities of the stators with other components of the BFM.

Low torque might underlie ATM in two ways. First, it affects tumbling. Because tumbling is influenced by the frequency of flagellar rotation amongst other factors, which is a function of torque generated by the BFM, it is also influenced by the torque [88]. Second, torque could be of importance in allowing the cells to pass through narrow pores — low torque may increase the likelihood with which cells get trapped, contributing to ATM as described above. Structural incompatibilities of the stators with other components of the BFM may be the ultimate cause of low torque.

2.4.4.2 ATM as early-stage phenotype in evolution of the BFM?

Although the underlying causes of ATM are unclear at this moment, it might be nice to speculate briefly about the evolutionary relevance of this type of motility, as it represents an improvement in comparison with a population of non-motile cells. As such, it might represent an early-stage intermediate in the evolution from a population of non-motile cells through a population of randomly motile cells [29] to a population of chemotaxis-performing cells. If we assume that ATM is a combination of above mentioned underlying possibilities, it might be that ATM is a result of slow-motile, chemotactic-deficient cells that harbour a BFM which provides a low torque. In the origin of the BFM, this might have been one of the first improvements from being non-motile. Addition of a new component might have caused the ancestral BFM to be able to generate a low torque. With an ancient type of filament this might have led to a cell that was poorly motile and non-chemotactic, as this is proposed to have evolved later [29].

2.5 Acknowledgements

We thank Ilja Westerlaken for technical assistance.

2.6 Materials and Methods

2.6.1 Strains and growth conditions

All ancestral strains used in the selection experiment are listed in Table 2.3. The knock-out strain, *Escherichia coli* K12 MG1655 Δ *motAB*, was engineered by replacing the stator genes *motAB* with the *rpsL-Neo* cassette [89] through homologous recombination. Next, we replaced the *rpsL* gene with *rpsL150* through recombination using Lambda Red [90–92]. *rpsL150* was obtained by PCR from Top10 cells.

Stator genes were amplified by PCR from genomic DNA (for donor strains see Table 2.3), using appropriate primers. Next, the PCR fragments were ligated into the *smal* site of the low-copy plasmid pBAD33, which contained a chloramphenicol-resistance gene [93]. Ligating the PCR fragments into the *smal* site, put the stator genes under control of an arabinose-inducible promoter, which allowed us to turn expression of stator genes on or off. The gene encoding the far-red fluorescent protein mKate2 [94] was amplified by PCR, using a forward primer that contained the sequence for the same arabinose-inducible promoter as was present on the plasmid. Next, the PCR-fragment was cloned into pBADTJ01 downstream of the stator genes using HindIII and Sall restriction sites. We chose a reverse orientation of the mKate2 construct, compared with the stator genes, to prevent read through. By heat-shock transformation, plasmids were transformed into heat-shock-competent Δ *motAB* cells. Strain pBAD33 harboured the pBAD33 plasmid without stator genes.

The strains listed in Table 2.3 were always grown overnight before use (O/N; 5 μ l from -80°C stock; 16-17h; 250rpm; 37°C ; orbit diameter 2.5cm) in 5 ml liquid LB medium (20 g/l LB powder) supplemented with 0.0025% chloramphenicol (w/v), unless stated otherwise. Furthermore, all strains, including the genotypes isolated from flares that were generated during the selection

experiment (section 0), were stored at -80°C in 31.1% glycerol (v/v) supplemented with 0.38% NaCl (w/v), unless mentioned otherwise.

Table 2.3. Ancestral strains used in this chapter.

Strain ID	Host strain	Stator genes	Plasmid ID	Donor strain stator genes
Strain pBAD33	<i>Escherichia coli</i> K12 str. MG1655 Δ motAB (short Δ motAB)	-	pBAD33	-
Strain 01	Δ motAB	<i>motAB</i>	pBADTI01	<i>Escherichia coli</i> K12 str. MG1655
Strain 01-mKate2	Δ motAB	<i>motAB</i>	pBADTI01-mKate2	<i>Escherichia coli</i> K12 str. MG1655
Strain 02	Δ motAB	<i>motAB</i>	pBADTI02	<i>Escherichia coli</i> O111:H-str. 11128
Strain 03	Δ motAB	<i>motAB</i>	pBADTI03	<i>Escherichia coli</i> O26:H11 str. 11368
Strain 05	Δ motAB	<i>motAB</i> "1"	pBADTI05	<i>Bdellovibrio bacteriovorus</i> HD100
Strain 08	Δ motAB	<i>lafTU</i>	pBADTI08	<i>Escherichia coli</i> O111:H-str. 11128
Strain 09	Δ motAB	<i>motAB</i>	pBADTI09	<i>Azospirillum</i> sp. B510
Strain 11	Δ motAB	<i>motPS</i>	pBADTI11	<i>Bacillus pseudofirmus</i> OF4
Strain 13	Δ motAB	<i>lafTU</i>	pBADTI13	<i>Photobacterium profundum</i> SS9
Strain 14	Δ motAB	<i>lafTU</i> "2"	pBADTI14	<i>Vibrio shilonii</i> AK1
Strain 15	Δ motAB	<i>pomAB</i>	pBADTI15	<i>Shewanella oneidensis</i> MR-1
Strain 16	Δ motAB	<i>motAB</i>	pBADTI16	<i>Bacillus megaterium</i> DSM319
Strain 17	Δ motAB	<i>motAB</i>	pBADTI17	<i>Listeria monocytogenes</i> EGD-e
Strain 18	Δ motAB	<i>motPS</i>	pBADTI18	<i>Bacillus megaterium</i> DSM319
Strain 19	Δ motAB	<i>motAB</i> Na putative	pBADTI19	<i>Photobacterium profundum</i> SS9
Strain 20	Δ motAB	<i>lafTU</i> "1"	pBADTI20	<i>Vibrio shilonii</i> AK1
Strain 21	Δ motAB	<i>pomAB</i>	pBADTI21	<i>Vibrio shilonii</i> AK1
Strain 22	Δ motAB	<i>motAB</i>	pBADTI22	<i>Edwardsiella ictaluri</i> 93-146
Strain 23	Δ motAB	<i>motAB</i>	pBADTI23	<i>Shewanella oneidensis</i> MR-1
Strain 24	Δ motAB	<i>motAB</i>	pBADTI24	<i>Proteus mirabilis</i> HI4320
Strain 25	Δ motAB	<i>motAB</i>	pBADTI25	<i>Pseudomonas aeruginosa</i> UCBPP-PA14
Strain 26	Δ motAB	<i>motAB</i>	pBADTI26	<i>Sodalis glossinidius</i> str. 'morsitans'
Strain 28	Δ motAB	<i>motCD</i>	pBADTI28	<i>Pseudomonas aeruginosa</i> UCBPP-PA14
Strain 29	Δ motAB	<i>motAB</i> "1"	pBADTI29	<i>Rhodospirillum centenum</i> SW

2.6.2 Selection experiment

5 μl from an O/N culture of an ancestral strain (Table 2.3) was inoculated in the centre of a semi-solid agar plate (90 mm diameter petri dish). Semi-solid agar plates consisted of 27 ml of 50% LB medium (v/v; Tryptone, 10 g/l; Yeast Extract, 2.5 g/l; NaCl, 10 g/l) supplemented with 0.3% agar (w/v), 0.2% L-arabinose (w/v) and 0.0025% chloramphenicol (w/v). For each strain, we started with 10 replicate populations. Semi-solid agar plates, with a mark on the side (for orientation; see below),

were poured 19-20h in advance (also in other experiments) and were dried statically at room temperature in stacks of 5 plates. After inoculation with cells, the semi-solid agar plates were incubated statically in stacks of 3 in airtight, plastic boxes (12 cm x 12 cm x 5cm) at 37°C. Every 24h (one round of selection), the semi-solid agar plates were checked for evolution of increased fitness, visible as flares. We defined flares as wedge-shaped populations of mutant cells that escaped from the population front of their direct ancestor. Three different outcomes after a round of selection were possible. 1) If the ancestral population was motile immediately and had reached the edge of the semi-solid agar plate without generation of a flare, then this marked the end of that selection line. 2) If the ancestral population was either non-motile without generation of a flare or motile without reaching the edge, then the population was incubated for another 24h. This was repeated for up to one week (no checks during weekends), after which a sample of the population was transferred to fresh semi-solid agar. 200 µl from the outer edge of the population was sampled with a cut-off 1000 µl pipette tip and diluted in 200 µl 70% glycerol (v/v) supplemented with 0.85% NaCl (w/v). Before storage of this sample at -150°C, 5ul was inoculated in the centre of a fresh semi-solid agar plate for additional selection for up to 4 weeks. 3) If evolution of increased fitness had occurred, the generated flares were selected according to pre-determined rules. If the ancestral population was non-motile, flares had to be ≥ 10 mm in radius before they were selected. If the ancestral population was motile, flares were allowed to have a smaller radius also. If several flares were generated on one semi-solid agar plate, they were selected in order of appearance to prevent selection bias. The semi-solid agar plate, with the side-mark pointing towards the experimenter, was turned clockwise and the first flare pointing towards the experimenter was selected. Next, the plate was turned further to select more flares. 200 µl was sampled from the rim of a flare (if a flare had reached the edge, the flare was sampled at the edge) with a cut-off 1000 µl pipette tip. This sample was stored immediately at -80°C in 200 µl 70% glycerol (v/v) supplemented with 0.85% NaCl (w/v). Before this sample was stored permanently at -150°C, it was defrosted and a few µl was streaked on LB agar plates containing 1% agar (w/v) and 0.0025% chloramphenicol (w/v). Plates were incubated statically at 37°C for up to 24h. A single colony per flare was selected for O/N growth and storage at -80°C. 5ul from this -80°C stock was then inoculated on a fresh semi-solid agar plate for additional selection.

2.6.3 Nomenclature of strains that evolved in selection experiment

During the selection experiment, we expected many flares to be generated. To be able to document all genotypes that we isolated from these flares (called strains) in an orderly fashion, we devised a nomenclature that is shown in Figure 2.10. All strains were assigned a number, which represented the rank of the flare it was isolated from, and a letter, which distinguished between multiple strains that evolved from a single direct ancestor. Primary flares were generated by the ancestral strain X, secondary flares were generated by genotypes isolated from primary flares, and so on. For instance, strain X.1a evolved from ancestral strain X, X.1a.2a evolved from X.1a, and X.1a.2a.3b evolved from X.1a.2a. Strains indicated by a red circle were called endpoints and represented strains that did not evolve improved motility during the selection experiment. Strains indicated by blue circles highlight an evolutionary lineage that connects an endpoint to the ancestor.

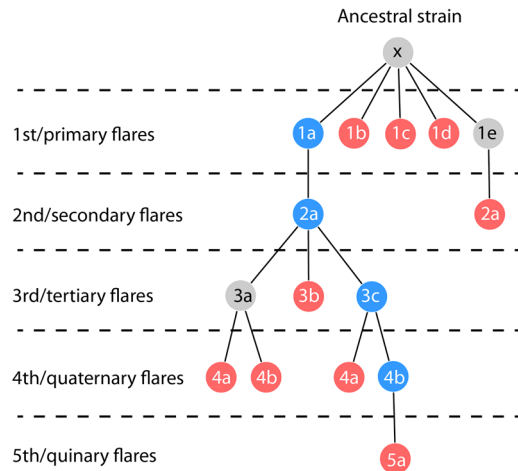


Figure 2.10. Nomenclature of strains isolated from flares that were generated in the selection experiment. The ancestral strain represents 10 replicate populations. Numbers (1, 2, 3, etc.) indicate the rank of the flare, letters (a, b, c, etc.) distinguish between strains that evolved from the same ancestor. Red strains represent endpoints, which are genotypes that were not able to evolve improved motility in the selection experiment. The blue strains highlight an evolutionary lineage of strains that connects an endpoint to the ancestor: $X \rightarrow X.1a \rightarrow X.1a.2a \rightarrow X.1a.2a.3c \rightarrow X.1a.2a.3c.4b \rightarrow X.1a.2a.3c.4b.5a$.

2.6.4 Relative population swimming speed

5 μ l from an O/N culture was inoculated in the centre of a semi-solid agar plate. The semi-solid agar plates were prepared as described before, except that the plates were dried in stacks of 3 plates. Next, the semi-solid agar plates were incubated undisturbed in stacks of 3 plates in airtight, plastic boxes at 37°C. From 4h onwards, the distance the front of the population had travelled was measured every hour. In between two measurements, the populations were re-incubated at 37°C. As the travelled distance increased linearly with time [30], the population swimming speed could be calculated as $\frac{\Delta distance}{\Delta time}$. Unless stated otherwise, three measurement points were used for calculating the population swimming speed. Relative population swimming speed of a focal strain was calculated as $\frac{Speed\ focal\ strain}{Speed\ Strain\ 01}$.

2.6.5 Competitive motility

O/N cultures of a focal strain and the reference strain that expressed a fluorescent protein (strain O1-mKate2) were diluted with liquid LB medium to an optical density at 600 nm (OD600) of 1.0. From both of these dilutions, 5 μ l was inoculated in the centre of a semi-solid agar plate (dried in stacks of 3 plates; one stack was used for one focal strain), with a distance of 1 cm in between both points of inoculation (Figure 2.11; red dots). Next, the semi-solid agar plates were incubated undisturbed for >24 hours in stacks of 3 plates in airtight, plastic boxes at 37°C. Initial experiments had shown before that the pattern of competition between two strains did not change after 24 hours and we propose

that it did not change after the competing strains reached the edge of the agar plate (data not shown).

In order to determine competitive motility of focal strains, we used a Typhoon laser scanner (GE Healthcare Life Sciences). We scanned 12 semi-solid agar plates at a time (full capacity; Figure 2.11A), of which three plates were control plates (no cells, strain 01, strain 01-mKate2). The other nine plates consisted of three times three replicates of one focal strain. The Typhoon settings were as follows: excitation/emission filters, 532 nm/610 nm; focal plane, +3mm; pixel size, 1000 μm ; PMT, 600V; sensitivity, high. For every plate with competing strains, we next determined the radius of the circle that would fit the interaction boundary between two competing strains. As was shown before for growing populations of bacteria [48], the ratio of expansion velocities could be calculated from the radius of this circle. Here we used the *segmented line*-function (Figure 2.11B and D) and the *fit circle*-function (Figure 2.11C and E) of ImageJ to fit a circle to a manually-created segmented line that consisted of three anchor points. In order to fit a circle with similar characteristics to vastly different interaction boundaries (e.g. Figure 2.11B and D), these anchor points were chosen so as to only use a small part of the interaction boundary. For every plate, five replicate circles were created. The average radius (in cm) of these five circles was used to calculate the selective advantage (s) of one strain over the other. Rewriting the formulas provided in [48], gave us $s = \left(\frac{0.6427}{\text{radius circle}}\right)^{\frac{1}{0.918}}$. Using s , the ratio of expansion velocities was calculated, which we called competitive motility. When strain 01-mKate2 was more motile than the focal strain (circle to left; Figure 2.11E) we used $\text{competitive motility} = \frac{1}{s+1}$. When the focal strain was more motile than strain 01-mKate2 (circle to right; Figure 2.11C), we used $\text{competitive motility} = s + 1$.

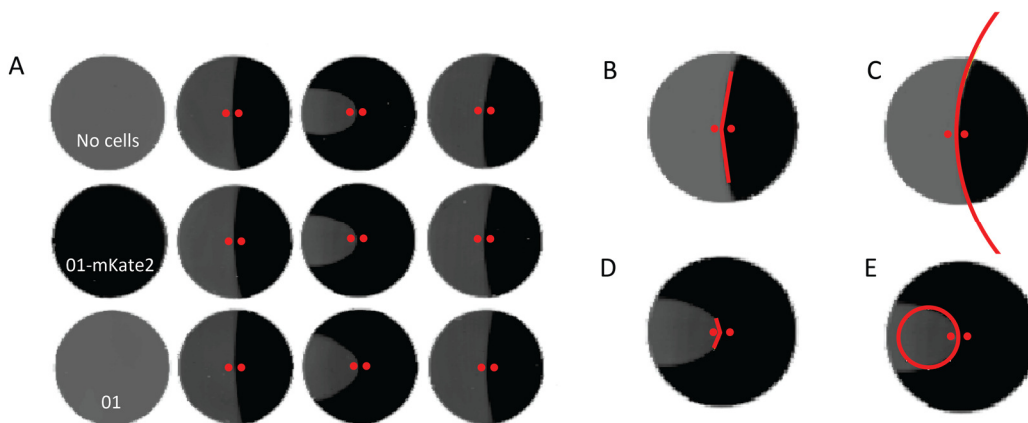


Figure 2.11. Typhoon images of competing strains on semi-solid agar plates. **A.** Overview of one scan consisting of 12 plates. The columns consisted of three replicate plates, except for the left column. On plates with competing strains the approximate positions of inoculation are indicated with red dots. The left dot represents the focal strain, the right dot represents strain 01-mKate2. **B and D.** Zoom-in on two plates with the segmented-line that is created with the *segmented line*-function of ImageJ. Three anchor points per line were used: begin, middle, end. **C and E.** Same images as B and D, respectively, but now with the circle that is created using the *fit circle*-function of ImageJ. In C only part of the circle is shown.

2.6.6 Single-cell measurements

A strain to be measured at the single-cell level was first grown O/N. Next, a sample of the culture was diluted 10x in liquid LB medium supplemented with 0.0025% chloramphenicol (w/v) and 0.2% L-arabinose (w/v). This culture was incubated for one hour at 37°C (250rpm, orbit diameter 2.5cm), after which a sample was washed twice by centrifugation and resuspension in liquid LB medium supplemented with 0.0025% chloramphenicol (w/v) and 0.2% L-arabinose (w/v). Optical density at 600 nm of this sample was adjusted to a value of 0.2-0.3, using the same medium, after which the sample was put on ice.

Next, this sample was added to a BSA-coated channel that was prepared one day in advance. A BSA-coated channel was ± 3 mm wide and ± 1 mm high and three channels were made at once using double-sided tape in between two 24x50 mm glass coverslips (Figure 2.12). After construction of the device, a solution of 0.5% BSA (w/v) was added to the channels in order to coat the glass surface. The channels were left to dry for 16-17h. Before adding a sample to the BSA-coated channel by pipetting, one end of the channel was sealed with vacuum silicone (Beckman Vacuum Grease Silicone). Once the channel was completely filled with sample, the other end was sealed with vacuum silicone as well. This was done to minimize the effects of convection. Measurements of single-cell phenotypes were performed immediately after filling the BSA-coated channel, with a temperature-controlled (37°C) widefield microscope (OKO) using a 20x objective. Each measurement was performed for 30 seconds at 8 frames per second (fps) and was saved as a TIFF stack. The recorded TIFF-stack was analysed by ImageJ MOSAIC ParticleTracker and the obtained data was loaded into a self-developed MATLAB script for further analysis [95].

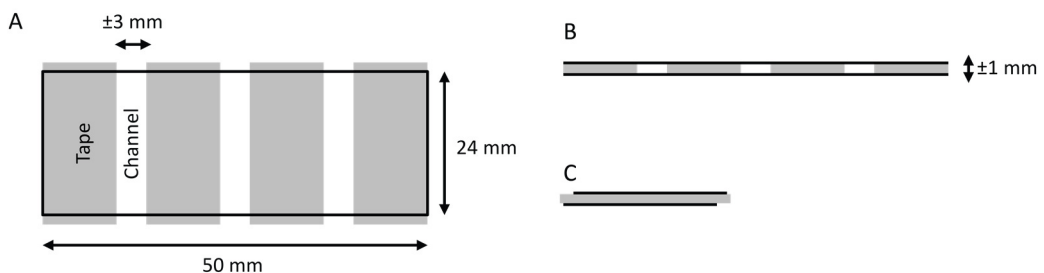


Figure 2.12. Device with three BSA-coated channels. **A.** Top view. The rectangle of 24x50 mm represents the coverslip. Grey rectangles represent double-sided tape. BSA-coated channels are indicated in white. **B.** Front view. The double-sided tape is sandwiched between two glass coverslips. **C.** Side view. The small overhang of the coverslips increases the efficiency of sealing the ends of the BSA flow channels with vacuum silicone.

2.6.7 Statistics

Different statistical analyses were performed in this chapter, depending on the distribution of the data and depending on the goal of the analysis. See text for details. When a correction for multiple comparisons was performed, we used the Benjamini-Hochberg procedure [96].

2.7 Literature

- 1 Gabaldón T, Rainey D & Huynen M (2005) Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I). *J. Mol. Biol.* **348**, 857–70.
- 2 Forgac M (2007) Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. *Nat. Rev. Mol. Cell Biol.* **8**, 917–29.
- 3 Finnigan GC, Hanson-Smith V, Stevens TH & Thornton JW (2012) Evolution of increased complexity in a molecular machine. *Nature* **481**, 360–4.
- 4 Sowa Y & Berry RM (2008) Bacterial flagellar motor. *Q. Rev. Biophys.* **41**, 103–32.
- 5 Pallen M & Matzke N (2006) From The Origin of Species to the origin of bacterial flagella. *Nat. Rev. Microbiol.* **4**, 784–790.
- 6 Liu R & Ochman H (2007) Stepwise formation of the bacterial flagellar system. *Proc. Natl. Acad. Sci.* **104**, 7116–7121.
- 7 Mulkidjanian A & Makarova K (2007) Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. *Nat. Rev. Microbiol.* **5**, 892–899.
- 8 Dolezal P, Likic V, Tachezy J & Lithgow T (2006) Evolution of the molecular machines for protein import into mitochondria. *Science* **313**, 314–8.
- 9 Clements A, Bursac D, Gatsos X, Perry AJ, Civciristov S, Celik N, Likic VA, Poggio S, Jacobs-Wagner C, Strugnell RA & Lithgow T (2009) The reducible complexity of a mitochondrial molecular machine. *Proc. Natl. Acad. Sci.* **106**, 15791–15795.
- 10 Archibald JM, Logsdon Jr. JM & Doolittle WF (2000) Origin and Evolution of Eukaryotic Chaperonins: Phylogenetic Evidence for Ancient Duplications in CCT Genes. *Mol. Biol. Evol.* **17**, 1456–1466.
- 11 Seidl MF & Schultz J (2009) Evolutionary flexibility of protein complexes. *BMC Evol. Biol.* **9**, 155.
- 12 Watson RA (2002) *Compositional Evolution: Interdisciplinary Investigations in Evolvability, Modularity and Symbiosis*.
- 13 Lind PA, Tobin C, Berg OG, Kurland CG & Andersson DI (2010) Compensatory gene amplification restores fitness after inter-species gene replacements. *Mol. Microbiol.* **75**, 1078–89.
- 14 Omer S, Kovacs A, Mazor Y & Gophna U (2010) Integration of a foreign gene into a native complex does not impair fitness in an experimental model of lateral gene transfer. *Mol. Biol. Evol.* **27**, 2441–5.
- 15 Perica T, Chothia C & Teichmann S a (2012) Evolution of oligomeric state through geometric coupling of protein interfaces. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8127–32.
- 16 Hashimoto K & Panchenko AR (2010) Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20352–20357.
- 17 Gray MW, Lukeš J, Archibald JM, Keeling PJ & Doolittle WF (2010) Irremediable Complexity? *Science* **330**, 920–921.
- 18 Wagner A (2011) The low cost of recombination in creating novel phenotypes: Recombination can create new phenotypes while disrupting well-adapted phenotypes much less than mutation. *BioEssays* **33**, 636–46.
- 19 Gophna U & Ofra Y (2011) Lateral acquisition of genes is affected by the friendliness of their products. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 343–8.

- 20 Watson RA & Pollack JB (2005) Modular Interdependency in Complex Dynamical Systems. *Artif. Life* **11**, 445–457.
- 21 Blair DF (2003) Flagellar movement driven by proton translocation. *FEBS Lett.* **545**, 86–95.
- 22 Chevance FF V & Hughes KT (2008) Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.* **6**, 455–65.
- 23 Zhou J, Sharp LL, Tang HL, Lloyd SA, Billings S, Braun TF & Blair DF (1998) Function of Protonatable Residues in the Flagellar Motor of *Escherichia coli*: a Critical Role for Asp 32 of MotB. *J. Bacteriol.* **180**, 2729–2735.
- 24 Zhou J & Blair DF (1997) Residues of the Cytoplasmic Domain of MotA Essential for Torque Generation in the Bacterial Flagellar Motor. *J. Mol. Biol.* **273**, 428–439.
- 25 Lloyd SA & Blair DF (1997) Charged Residues of the Rotor Protein FliG Essential for Torque Generation in the Flagellar Motor of *Escherichia coli*. *J. Mol. Biol.* **266**, 733–744.
- 26 Zhou J, Lloyd SA & Blair DF (1998) Electrostatic interactions between rotor and stator in the bacterial flagellar motor. *Proc. Natl. Acad. Sci.* **95**, 6436–6441.
- 27 Berg HC & Anderson RA (1973) Bacteria Swim by Rotating their Flagellar Filaments. *Nature* **245**, 380–382.
- 28 Larsen SH, Reader RW, Kort EN, Tso W-W & Adler J (1974) Change in direction of flagellar rotation is the basis of the chemotactic response in *Escherichia coli*. *Nature* **249**, 74–77.
- 29 Wei Y, Wang X, Liu J, Nememan I, Singh AH, Weiss H & Levin BR (2011) The population dynamics of bacteria in physically structured habitats and the adaptive virtue of random motility. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4047–52.
- 30 Adler J (1966) Chemotaxis in bacteria. *Science* **153**, 708–716.
- 31 Adler J (1966) Effect of amino acids and oxygen on chemotaxis in *Escherichia coli*. *J. Bacteriol.* **92**, 121–129.
- 32 Berg HC & Brown DA (1972) Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature* **239**, 500–504.
- 33 Wadhams GH & Armitage JP (2004) Making sense of it all: bacterial chemotaxis. *Nat. Rev. Mol. Cell Biol.* **5**, 1024–37.
- 34 Welch M, Oosawat K & Aizawat S (1993) Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. *Proc. Natl. Acad. Sci.* **90**, 8787–8791.
- 35 Toker A & Macnab R (1997) Distinct regions of bacterial flagellar switch protein FliM interact with FliG, FliN and CheY. *J. Mol. Biol.* **273**, 623–34.
- 36 Chun S & Parkinson J (1988) Bacterial motility: membrane topology of the *Escherichia coli* MotB protein. *Science* **239**, 276–278.
- 37 Khan S, Dapice M & Reese TS (1988) Effects of mot Gene Expression on the Structure of the Flagellar Motor at the Marine Biological Laboratory. *J. Mol. Biol.* **202**, 575–584.
- 38 Paulick A, Koerdt A, Lassak J, Huntley S, Wilms I, Narberhaus F & Thormann KM (2009) Two different stator systems drive a single polar flagellum in *Shewanella oneidensis* MR-1. *Mol. Microbiol.* **71**, 836–50.
- 39 Asai Y, Yakushi T, Kawagishi I & Homma M (2003) Ion-coupling Determinants of Na⁺-driven and H⁺-driven Flagellar Motors. *J. Mol. Biol.* **327**, 453–463.
- 40 Gosink KK & Ha CC (2000) Requirements for Conversion of the Na⁺-Driven Flagellar Motor of *Vibrio cholerae* to the

H⁺-Driven Motor of *Escherichia coli*. *J. Bacteriol.* **182**, 4234–4240.

- 41 Garza A, Harris-Haller L, Stoebner R & Manson M (1995) Motility protein interactions in the bacterial flagellar motor. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 1970–4.
- 42 Garza A, Biran R, Wohlschlegel J & Manson M (1996) Mutations in *motB* suppressible by changes in stator or rotor components of the bacterial flagellar motor. *J. Mol. Biol.* **258**, 270–85.
- 43 Marykwas DL & Berg HC (1996) A mutational analysis of the interaction between FliG and FliM, two components of the flagellar motor of *Escherichia coli*. *J. Bacteriol.* **178**, 1289–94.
- 44 Passmore SE, Meas R & Marykwas DL (2008) Analysis of the FliM/FliG motor protein interaction by two-hybrid mutation suppression analysis. *Microbiology* **154**, 714–24.
- 45 Hosking ER & Manson MD (2008) Clusters of charged residues at the C terminus of MotA and N terminus of MotB are important for function of the *Escherichia coli* flagellar motor. *J. Bacteriol.* **190**, 5517–21.
- 46 Hallatschek O, Hersen P, Ramanathan S & Nelson DR (2007) Genetic drift at expanding frontiers promotes gene segregation. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 19926–30.
- 47 Hallatschek O & Nelson DR (2010) Life at the front of an expanding population. *Evolution* **64**, 193–206.
- 48 Korolev K, Müller M, Karahan N, Murray A, Hallatschek O & Nelson D (2012) Selective sweeps in growing microbial colonies. *Phys. Biol.* **9**, 26008.
- 49 Toh H, Weiss BL, Perkin S a H, Yamashita A, Oshima K, Hattori M & Aksoy S (2006) Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res.* **16**, 149–56.
- 50 Koster DA, Mayo A, Bren A & Alon U (2012) Surface growth of a motile bacterial population resembles growth in a chemostat. *J. Mol. Biol.* **424**, 180–91.
- 51 van Vliet S, Hol FJH, Weenink T, Galajda P & Keymer JE (2014) The effects of chemical interactions and culture history on the colonization of structured habitats by competing bacterial populations. *BMC Microbiol.* **14**, 116.
- 52 Pigliucci M (2005) Evolution of phenotypic plasticity: where are we going now? *Trends Ecol. Evol.* **20**, 481–6.
- 53 Vladimirov N, Løvdok L, Lebedz D & Sourjik V (2008) Dependence of bacterial chemotaxis on gradient shape and adaptation rate. *PLoS Comput. Biol.* **4**, e1000242.
- 54 Berg HC (2003) The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19–54.
- 55 Berg HC (2004) *E. coli in motion* Springer.
- 56 Armstrong JB & Adler J (1969) Complementation of nonchemotactic mutants of *Escherichia coli*. *Genetics* **61**, 61–66.
- 57 Braun TF, Poulson S, Gully JB, Empey JC, Van Way S, Putnam a & Blair DF (1999) Function of proline residues of MotA in torque generation by the flagellar motor of *Escherichia coli*. *J. Bacteriol.* **181**, 3542–51.
- 58 Brown PN, Terrazas M, Paul K & Blair DF (2007) Mutational analysis of the flagellar protein FliG: sites of interaction with FliM and implications for organization of the switch complex. *J. Bacteriol.* **189**, 305–12.
- 59 Fujinami S, Terahara N, Lee S & Ito M (2007) Na⁺ and flagella-dependent swimming of alkaliphilic *Bacillus pseudofirmus* OF4: a basis for poor motility at low pH and enhancement in viscous media in an “up-motile” variant. *Arch. Microbiol.* **187**, 239–47.

- 60 Terahara N, Sano M & Ito M (2012) A *Bacillus* flagellar motor that can use both Na⁺ and K⁺ as a coupling ion is converted by a single mutation to use only Na⁺. *PLoS One* **7**, e46248.
- 61 Ito M, Hicks DB, Henkin TM, Guffanti AA, Powers BD, Zvi L, Uematsu K & Krulwich TA (2004) MotPS is the stator-force generator for motility of alkaliphilic *Bacillus*, and its homologue is a second functional Mot in *Bacillus subtilis*. *Mol. Microbiol.* **53**, 1035–49.
- 62 Lenski RE, Rose MR, Simpson SC, Tadler SC, Naturalist TA & Dec N (1991) Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2,000 Generations. *Am. Nat.* **138**, 1315–1341.
- 63 Rainey PB & Travisano M (1998) Adaptive radiation in a heterogeneous environment. *Nature* **32**, 69–72.
- 64 Poelwijk FJ, Kiviet DJ, Weinreich DM & Tans SJ (2007) Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383–6.
- 65 Orr HA (2003) The Distribution of Fitness Effects Among Beneficial Mutations. *Genetics* **163**, 1519–1526.
- 66 Moore FB, Rozen DE & Lenski RE (2000) Pervasive compensatory adaptation in *Escherichia coli*. *Proc. Biol. Sci.* **267**, 515–22.
- 67 Hall AR, Griffiths VF, MacLean RC & Colegrave N (2010) Mutational neighbourhood and mutation supply rate constrain adaptation in *Pseudomonas aeruginosa*. *Proc. Biol. Sci.* **277**, 643–50.
- 68 Chen S, Beeby M, Murphy GE, Leadbetter JR, Hendrixson DR, Briegel A, Li Z, Shi J, Tocheva EI, Müller A, Dobro MJ & Jensen GJ (2011) Structural diversity of bacterial flagellar motors. *EMBO J.* **30**, 2972–81.
- 69 de Mot R & Vanderleyden J (1994) The C-terminal sequence conservation between OmpA-related outer membrane proteins and MotB suggests a common function in both Gram-positive and Gram-negative bacteria, possibly in the interaction of these domains with peptidoglycan. *Mol. Microbiol.* **12**, 333–334.
- 70 Koebnik R (1995) Proposal for a peptidoglycan-associating alpha-helical motif in the C-terminal regions of some bacterial cell-surface proteins. *Mol. Microbiol.* **16**, 1269–1270.
- 71 Ma J, Campbell A & Karlin S (2002) Correlations between Shine-Dalgarno Sequences and Gene Features Such as Predicted Expression Levels and Operon Structures. *J. Bacteriol.* **184**, 5733–5745.
- 72 Plotkin JB & Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42.
- 73 Goodman DB, Church GM & Kosuri S (2013) Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science* **342**, 475–479.
- 74 Gophna U, Ron EZ & Graur D (2003) Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene* **312**, 151–163.
- 75 Park JY, Yoo SJ, Hwang CM & Lee S-H (2009) Simultaneous generation of chemical concentration and mechanical shear stress gradients using microfluidic osmotic flow comparable to interstitial flow. *Lab Chip* **9**, 2194–202.
- 76 Righetti P, Brost B & Snyder R (1981) On the limiting pore size of hydrophilic gels for electrophoresis and isoelectric focusing. *J. Biochem. Biophys. Methods* **4**, 347–363.
- 77 Pernodet N, Tinland B, Sadron IC & Pasteur CL (1997) Pore size of agarose gels by atomic force microscopy. *Electrophoresis* **18**, 55–58.
- 78 Narayanan J, Xiong J-Y & Liu X-Y (2006) Determination of agarose gel pore size: Absorbance measurements vis a vis

other techniques. *J. Phys. Conf. Ser.* **28**, 83–86.

79 Croze OA, Ferguson GP, Cates ME & Poon WCK (2011) Migration of chemotactic bacteria in soft agar: role of gel concentration. *Biophys. J.* **101**, 525–34.

80 Adler J (1975) Chemotaxis in bacteria. *Annu. Rev. Biochem.* **44**, 341–356.

81 Wolfe AJ & Berg HC (1989) Migration of bacteria in semisolid agar. *Proc. Natl. Acad. Sci.* **86**, 6973–6977.

82 Behe MJ (2001) Reply to My Critics: A Response to Reviews of Darwin's Black Box : The Biochemical Challenge to Evolution. *Biol. Philos.* **16**, 685–709.

83 Rorick M (2012) Quantifying protein modularity and evolvability: a comparison of different techniques. *Biosystems.* **110**, 22–33.

84 Wagner GP, Pavlicev M & Cheverud JM (2007) The road to modularity. *Nat. Rev. Genet.* **8**, 921–31.

85 Cooper V & Lenski R (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* **407**, 736–9.

86 de Visser JAGM & Lenski RE (2002) Long-term experimental evolution in *Escherichia coli*. XI. Rejection of non-transitive interactions as cause of declining rate of adaptation. *BMC Evol. Biol.* **2**, 19.

87 Buckling A, Wills M a & Colegrave N (2003) Adaptation limits diversification of experimental bacterial populations. *Science* **302**, 2107–9.

88 Bai F, Minamino T, Wu Z, Namba K & Xing J (2012) Coupling between switching regulation and torque generation in bacterial flagellar motor. *Phys. Rev. Lett.* **108**, 178105.

89 Heermann R, Zeppenfeld T & Jung K (2008) Simple generation of site-directed point mutations in the *Escherichia coli* chromosome using Red(R)/ET(R) Recombination. *Microb. Cell Fact.* **7**, 14.

90 Datsenko KA & Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–5.

91 Mosberg J, Lajoie M & Church G (2010) Lambda red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate. *Genetics* **186**, 791–9.

92 Thomason LC, Sawitzke JA, Li X, Costantino N & Court DL (2014) Recombineering: genetic engineering in bacteria using homologous recombination. *Curr. Protoc. Mol. Biol.* **106**, 1.16.1-1.16.39.

93 Guzman LM, Belin D, Carson MJ, Beckwith J, Guzman L, Belin D & Carson MJ (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130.

94 Shcherbo D, Murphy CS, Ermakova G V, Solovieva E a, Chepurnykh T V, Shcheglov AS, Verkhusha V V, Pletnev VZ, Hazelwood KL, Roche PM, Lukyanov S, Zaraisky AG, Davidson MW & Chudakov DM (2009) Far-red fluorescent tags for protein imaging in living tissues. *Biochem. J.* **418**, 567–74.

95 Langelaan M (2014) Evolving the bacterial flagellar motor by horizontal gene transfer. .

96 Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300.

2.8 Supplementary information

2.8.1 Supplementary information A

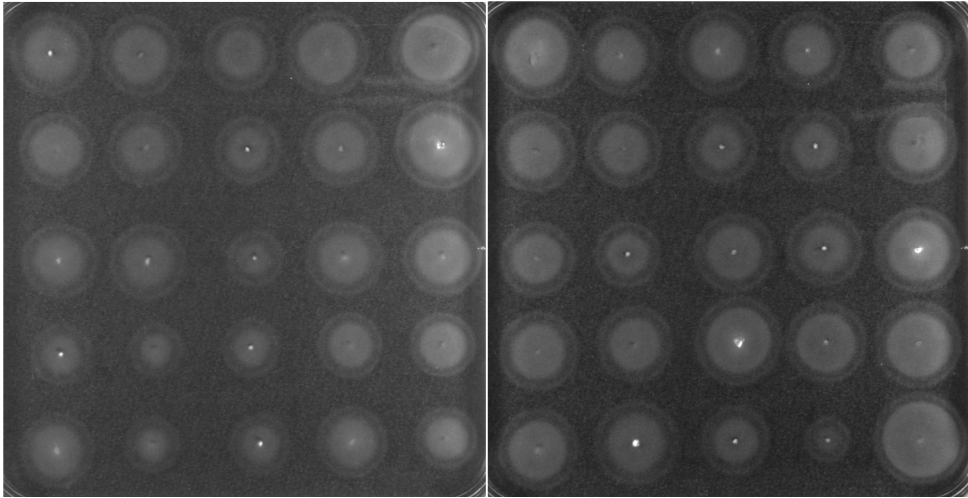


Figure 2.13. 50 genotypes isolated from a flare of strain 17. All isolated genotypes were more motile than the direct ancestor, which showed ATM. Differences in population diameter might be a result of different mutations or differences in population size of the inoculum.

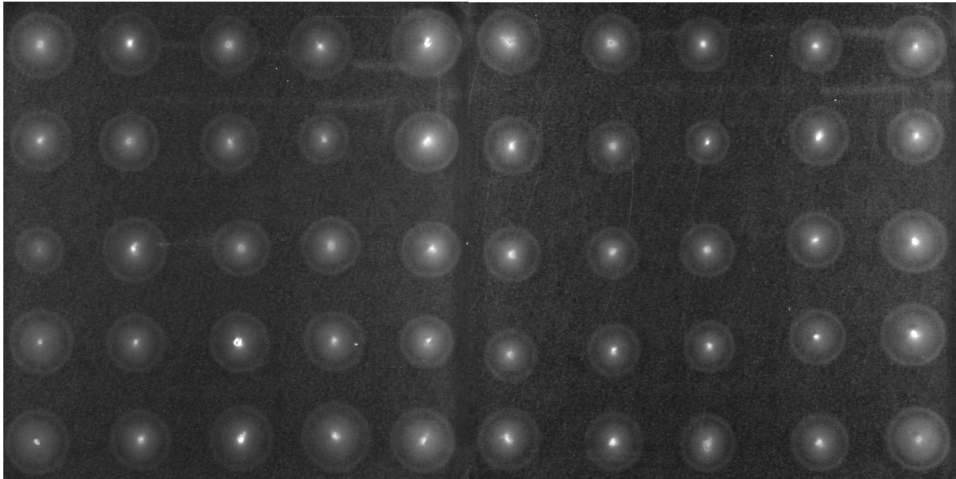


Figure 2.14. 50 genotypes isolated from another flare of strain 17. All isolated genotypes were more motile than the direct ancestor, which showed ATM. Differences in population diameter might be a result of different mutations or differences in population size of the inoculum.

2

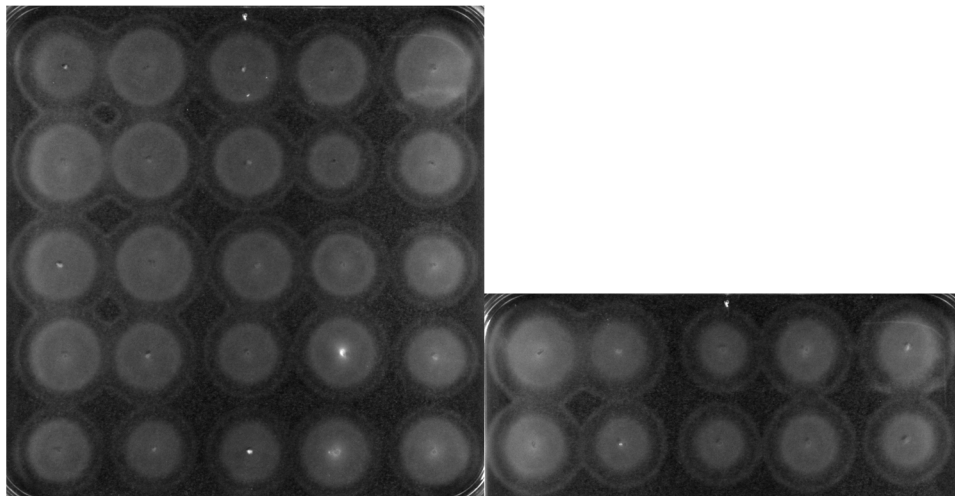


Figure 2.15. Genotypes isolated from a population of cells harbouring WT stators. Differences in population diameter might be a result of differences in population size of the inoculum.

3

The mutations underlying compositional evolution

Régis C.E. Flohr, Erwin van Rijn, Hubertus J.E. Beaumont. *In preparation for publication.*

3.1 Abstract

In the previous chapter, we described evolutionary integration of incompatible foreign stators in the bacterial flagellar motor (BFM) and other cell systems of *Escherichia coli*. This demonstrated experimentally the creative potential of evolution to resolve inter component conflicts that otherwise would limit compositional evolution. In this chapter we describe the results from genome sequence analyses aimed at revealing the entire set of mutations that facilitated this process. The results show that approximately 40% of the underlying mutations in structural genes clustered, especially in the foreign stator proteins and BFM proteins in their structural vicinity. The remaining mutations affected more distant BFM proteins but also a range of other cell systems that are structurally involved in the BFM's functionality. Strains with different incompatible stators followed overlapping but different genetic trajectories. In some cases, replicate lineages of strains expressing identical incompatible stators also followed divergent mutational trajectories. Together, these findings open a first window on the availability of mutational trajectories that facilitate compositional evolution and suggest that, at least in the case of our model, there are multiple ways to integrate a single foreign component.

3.2 Introduction

Experimental data about the underlying molecular mechanisms of protein-complex evolution is scarce, which is assumed to have occurred by the step-wise addition of pre-existing protein components to simpler assemblies [1–8]. Because these pre-existing protein components were not adapted to the initial assembly (unless it involved duplicates of components already being part of the complex [9]), it is likely that they did not interact perfectly with the protein complex upon addition. Furthermore, when new components are acquired from other organisms, e.g. through horizontal gene transfer (HGT), translational problems due to differences in codon usage and GC-content might occur [10,11]. In other words, fitness-increasing mutations are likely to occur upon addition of a foreign component to a protein complex. In this chapter, we will provide a first analysis of mutations underlying foreign component integration, which we described in the previous chapter.

The model system that we used to investigate the molecular mechanisms behind protein complex evolution, is the extensively studied bacterial flagellar motor (BFM) of *Escherichia coli* (*E. coli*; for reviews see [12–14]). Briefly, the BFM couples ion translocation to torque generation [12], is composed of more than 20 different protein components [12] and can be decomposed into four major structural features: the rotating basal body, the stator complex, the flagellar filament and the hook that connects the basal body and flagellar filament [14]. The BFM is essential for bacterial swimming and together with the chemotaxis signalling pathway [15], it allows cells to respond to chemical gradients in the environment [16]. Our main focus was on the stator complex of the BFM. The stator complex is composed of two proteins, MotA and MotB [17,18], and functions in anchoring the BFM to the cell wall [19,20], ion translocation [21–23] and flagellar rotation/torque generation through interaction with the basal body [12,24–26]. Without the stator complex, the BFM does not function [27,28].

Previously (Chapter 2), we replaced stator genes of *E. coli* MG1655 K-12 with stator genes from 22 other bacteria from across the bacterial kingdom, which resulted in chimeric BFMs (cBFMs). Thus, the resulting 22 different strains (cBFM strains) each harboured one set of foreign stator genes. Five of the cBFM strains contained compatible stators, determined by the observation of motility, and were not able to evolve improved motility in a selection experiment in nutrient-rich, semi-solid agar. Three additional cBFM strains also contained compatible stators, but these strains evolved improved motility and thus showed *evolutionary integration* of the foreign components. The remaining 14 cBFM strains contained incompatible stators initially, determined by the absence of motility, of which three strains evolved compatibility followed by the evolution of improved motility. Together, six cBFM strains showed evolutionary integration of a foreign component during the selection experiment.

Because phenotypic data of replicate selection lines suggested that different mutational trajectories were taken within each strain, one of the questions that we posed was whether we could detect these mutational differences between replicate selection lines. Answering this question additionally would shed light on the complexity of the initial problem: multiple mutational ways would indicate that the evolutionary problem was less complex compared to the case when only one solution would exist.

Another question we hoped to answer was whether we could detect mutational differences between cBFM strains. Six foreign stators from different species of bacteria were evolutionarily

integrated, and it would be of interest to know the degree to which these required similar or different modifications. If mutational differences were found, this would provide insight into to diversity of homologous components with similar functions.

The third question we hoped to answer deals with the clustering of motility-increasing mutations. Although experimental evidence on the position of mutations that cause evolutionary integration is scarce [29–31], more experimental data exists about a related type of mutation that increases organismal fitness after deleterious mutations: compensatory mutations. For instance, it was shown that compensatory mutations are more likely to occur near the site of the deleterious mutation in terms of protein primary structure (and was proposed for protein tertiary structure) [32], which indicated that mutations were clustered. Another study [33] showed that compensatory mutations tended to be more effective when found closer, in terms of primary and tertiary structure, to the site of the deleterious mutation. Furthermore, in the BFM, deleterious mutations in stator protein MotB are known to be compensated for by mutations in either stator protein MotA or MotB or in switch-complex proteins FliG or FliM [34–36]. Also, compensatory mutations in MotA and MotB were discovered for deleterious mutations in MotA [36] and deleterious mutations in FliG were compensated for by mutations in FliM [37]. Together, these studies showed that fitness-increasing mutations following a deleterious mutation might be located to proteins outside of the initially-affected protein also. If these results were to be true for mutations that improve fitness after the addition of a foreign protein component to a protein complex, we would expect to find clustering of mutations, i.e. many mutations within the foreign protein components as well as in proteins in the structural vicinity of the stators.

In order to find the mutations underlying evolutionary integration, whole-genome re-sequencing by Next Generation Sequencing (NGS; by Illumina) was performed (*Materials and Methods*). With the much lower costs to perform whole-genome re-sequencing compared with a decade ago, NGS can be applied nowadays by individual labs to study, amongst others, microbial evolution in great detail (for review see [38]). The genomes (including plasmid DNA that contain the foreign stator genes) from endpoints, i.e. evolved genotypes that were not able to evolve improved motility during the selection experiment, from replicate selection lines of one strain were pooled, which allowed us to discover strain-specific mutations.

Our results showed that mutations, which resulted in the evolutionary integration of a foreign component, were clustered in all cBFM strains, especially around the foreign stator proteins. Besides strain-specific differences, we also identified differences in the selectively-accessible, mutational trajectories within a single cBFM strain. This result indicated that evolution found multiple ways to integrate a foreign component.

3.3 Results

Six cBFM strains evolved (improved) motility during the selection experiment. Of each strain, up to 14 genomes of endpoints from replicate selection lines were pooled. In total we made 10 pools, meaning that genotypes of strains with more than 14 endpoints were divided over multiple pools. This scheme allowed us make between-strain comparisons, but, for some strains, also within-strain comparisons. Furthermore, all 10 pools were analysed on the same flow cell, which reduced sample-

to-sample variation. Before analysing the sequencing results, we performed a preliminary filtering based on the quality (Q-scores) of the mutations that were discovered (*Materials and Methods*). Q-scores measure the probability that bases are called incorrectly. As the Q-scores determine the false positive rate (the higher the Q-score, the lower the false positive rate), we decided to only investigate mutations that had an average quality of ≥ 30 , which corresponds to a false positive rate of less than 0.1%. A Q-score of 30 is considered a benchmark in NGS, where virtually all reads will be perfect [39].

3.3.1 Validation of mutations discovered by Illumina NGS

In order to examine the accuracy with which our NGS approach successfully found mutations, we performed a number of validations with other means of mutation detection. We checked the presence of a selection of mutations that had been discovered by Illumina NGS sequencing in 4 of the 10 pools (Table 3.2). We focussed on mutations in genes of the BFM and genes of the chemotaxis signalling pathway, as we expected these genes to be likely candidates in evolutionary integration of the foreign stators. Furthermore, we would later examine some of these mutations in more detail when investigating mutational trajectories (mutations in cBFM strain 08 (pool 1) and 18 (pool 6); Chapter 4) and when investigating mutations that result in motility (mutations in cBFM strain 29 (pool 9 and 10); Chapter 5). The other mutations in strain 29 (FlIG I133L and E224Q) were selected in order to co-validate our own NGS analysis-software Antonie (these mutations were not found by Antonie; <http://bertusbeaumontlab.tudelft.nl/software.html>).

To examine the presence of the selected mutations, we used a combination of High Resolution Melt Analysis (HRMA; *Materials and Methods*) and Sanger sequencing (Table 3.1). Because we checked mutations with a large variety in measured characteristics (*Q-scores*, *read depths* and *presence mutations*), we assumed that the results obtained here could be extrapolated for all mutations found by NGS that were left after preliminary filtering. Although 3.8% of the mutations left after filtering had Q-scores between 30 and 36, and thus lied below the lowest Q-score of the mutations checked by other means of mutations detection, mutations with a Q-score in this range were assumed to yield similar results in terms of actual presence [39]. As mentioned above, *Q-score* (third column in Table 3.1) measures the probability that bases are called incorrectly. *Read depth* (fourth column in Table 3.1) indicates the number of NGS reads that contained the particular position (e.g. 212 NGS reads of strain 08 contained the nucleotide positions that encode amino acid position 280 in FlIG) and *presence mutation* (fifth column in Table 3.1) indicates the percentage of *read depth* that contains the particular mutation. As we sequenced a pool of up to 14 genomes, *presence mutation* depended on the number of different endpoints that contained the mutation.

The presence of the 23 mutations we checked ranged from 100% (all NGS reads in the pool of genomes contained the mutated base at a particular position) to 1.9% (98.1% of the NGS reads in the pool of genomes contained the original base at a particular position). Similarly, the Q-scores ranged from 36 to 255 and the read depth ranged from 182 to 7897. Of these mutations, 22 mutations were also found by HRMA and/or Sanger sequencing. With this fraction (22/23 or 95.7%), we estimated the 95% confidence interval using the modified Wald method [40] and found the interval ranging from 77.3% to 99.9%. In other words, based on these results, we estimated with 95% confidence that at least 77.3% of the mutations identified with our NGS approach were present in the evolved

genotypes. This implies that, if we identify two or more mutations in one gene, we can conclude with 95% confidence that this gene is mutated.

According to Illumina, their NGS sequencing could detect insertions and deletions (InDels) up to three basepairs (bp). In accordance with this estimate, we found that large InDels (> 8bp) were not discovered by the NGS filter as applied here (Chapter 5).

Table 3.1 Mutations found by NGS and checked for their presence by HRMA and/or Sanger sequencing. 4 out of 10 pools were used in this analysis. *Q-score* measures the probability that bases are called incorrectly. *Read depth* indicates the number of NGS reads that contained the bases encoding a particular amino acid. *Presence mutation* indicates the percentage of *read depth* that contains the particular mutation. *Pos* indicates the position of the mutated base on the plasmid. N.a.: not available.

Pool	Mutation	Q-score	Read depth	Presence mutation (%)	Found by HRM	Found by Sanger
1	LaFT E87K	255	7897	19.03	yes	yes
	LaFT Q90K	255	7850	8.95	yes	yes
	LaFT A134S	255	7643	99.99	yes	yes
	LaFT L284R	255	7340	35.54	yes	yes
	FlIG Q280R	68	212	10.38	yes	yes
6	MotA E91K	255	2638	6.52	yes	n.a.
	MotA G123R	255	2641	5.68	yes	yes
	MotA E125K	255	2633	19.26	yes	yes
	MotA L234R	123	2639	1.86	yes	yes
	MotA Q237K	255	2405	40.42	yes	yes
	MotB F29L	249	2659	3.46	yes	yes
	FlIC L10F	83	337	8.01	yes	yes
	FlIG A171T	52	237	7.17	yes	n.a.
	FlIG Q280R	68	268	8.21	yes	yes
	FlIM A161V	94	230	13.04	yes	n.a.
CheB V311L	151	256	18.36	yes	n.a.	
9	pos 6966	255	3941	10.71	yes	yes
	FlIG I133L	36	182	6.59	yes	yes
10	pos 6895	255	3928	13.47	yes	yes
	pos6931	255	3959	8.69	yes	yes
	pos 6964	255	3905	18.28	yes	yes
	pos 6966	255	3940	7.41	yes	yes
	FlIG E224Q	75	224	10.71	no	no

3.3.2 Genome-wide pattern of mutations in all evolved strains

In this section, we first focus on the overall patterns of mutations and do not distinguish between different strains. We intended to answer the following questions. How many genes were mutated with how many unique mutations, or in other words, how diverse were the genetic possibilities for evolutionary integration of the foreign stators? How did evolution affect protein function as opposed to protein regulation? Dividing the structural genes in subgroups, how clustered were mutations distributed? We focussed on genes of the BFM (*BFM genes*), genes involved in the chemotaxis signalling pathway (*chemotaxis genes*), genes that might affect the proton motive force (*PMF genes*), genes that might affect the peptidoglycan layer (*peptidoglycan genes*) and all other genes (*other genes*). Within the BFM and chemotaxis genes we also identified specific genes.

Our highest level of analysis examined the distribution of unique mutations between coding and non-coding (intergenic) regions of the genome. One unresolved question in evolutionary biology is whether genetic changes underlying evolution typically affect protein function or protein expression [41–44]. Because regulatory sequences (i.e. promoters) are often located in non-coding regions in the bacterial genome [45], investigating this distribution provides additional insight into this issue. As is shown in Figure 3.1 at the top left, 250 unique mutations were gained during the evolution experiment, of which 193 were located in 87 unique coding regions (genes). The remaining 57 mutations were found in non-coding (intergenic) regions, of which 3 might be located in the Shine-Dalgarno (SD) sequence of 3 different genes (*motA*, *sspA* and *flgB*; more detailed information on specific genes can be found in Table 3.3 to Table 3.8 in the supplementary information). The vast majority of the mutations comprised single nucleotide polymorphisms (SNPs; 238), with only a few insertions (Ins; 3) and deletions (Del; 9) accompanying the SNPs.

Next, we focussed on mutations in the coding regions only, making a distinction between regulatory and structural genes. We defined regulatory genes as genes of which the protein products are involved, or are predicted to be involved, in regulatory processes, such as regulating transcription and translation, thereby being at the basis of mRNA and protein availability. Structural genes were defined as genes of which the protein products are involved, or are predicted to be involved, in all other processes. Similar to the distribution of mutations between coding and non-coding regions of the DNA, our data might provide additional insight into the relative contribution of changes in protein function as opposed to changes in protein expression. We found 155 unique mutations in 69 structural genes, 33 mutations in 18 regulatory genes, and 5 mutations in 5 genes of which the function is not yet known or could not be predicted on the basis of similarity with other genes (Figure 3.1 top right). Although a significant number of mutations was found in regulatory genes, the distribution of mutations between regulatory and structural genes was in favour of structural genes.

In the class of structural genes, we defined five sub-classes: BFM genes, chemotaxis genes, PMF genes, peptidoglycan genes and other genes. We focussed on BFM and chemotaxis genes (together called *focal genes*), because they are known to have direct effects on motility in semi-solid agar [13]. Mutations in PMF genes were considered, based on the fact that they could have a rather direct influence on motility. Because the flow of ions through the stator causes rotation of the rotor and thus motility [21], changing the PMF might be a general way of influencing motility. PMF genes included genes that affected cellular growth and metabolism and could thus have affected fitness by other means than changing motility. Peptidoglycan genes were considered because of their possible influence on the positioning of the stator around the rotor [35]; MotB anchors to the peptidoglycan layer upon binding in the BFM [19,20,46,47]. The results (Figure 3.1 centre) showed that 70 unique mutations were located in 7 BFM genes with an additional 10 mutations in 4 chemotaxis genes, which resulted in 52% of the mutations in structural genes to be located in the focal genes. Taking also the mutations in PMF genes (18) and peptidoglycan genes (6) into account, we found 67% of the mutations in structural genes to be located in classes of genes we expected to be involved closely in the evolution of (improved) motility on semi-solid agar. The remaining 51 mutations (33% of mutations in structural genes) in 39 structural genes were involved in a plethora of processes, including amino acid synthesis, DNA recombination and membrane synthesis.

At the lowest level we examined the distribution of mutations among the *focal genes* that we expected to have major effects in the evolution of motility. Figure 3.1 (lower right) showed that 39 unique mutations (56% of the mutations in BFM genes) were located in the stator genes *motA* and *motB* that were foreign to the cell. An additional 14 mutations were found in *fliG*, of which the protein product is an interaction partner of MotA [12,13], and 11 mutations were found in *fliM*, of which the protein product is an interaction partner of FliG and CheY, the signalling component of the chemotaxis signalling pathway [12,13]. In total we found that 91% of the mutations in BFM genes was located in, or in close structural vicinity of, the stator proteins. An additional six mutations in three genes comprised the rest of the mutations in BFM genes, all of which code for protein products located further away from the stator proteins than FliG and FliM. FliC is the protein that forms the flagellar filament [48], FliH is part of the flagellar export apparatus [49], and FlgF is a component of the cell-proximal portion of the rod [49].

Mutations in the chemotaxis signalling pathway (Figure 3.1 lower left) were all located in four genes of which the protein products are not involved in direct interactions with the BFM. Rather, they are involved in the initial stages of signal transduction [15]. We found mutations in genes of the receptor (*tsr*), the modulator of the receptor (*cheB*), the modulator of the effector of the signalling pathway (*cheA*) and the connector between Tsr and CheA (*cheW*) [15].



Figure 3.1. Pattern of genome-wide unique mutations in end-points of all six cBFM strains that may underlie evolution of (improved) motility. The size of pie-chart sections represents the number of mutations.

At the beginning of this section we asked three questions. First, we were interested in the diversity of genetic possibilities for evolutionary integration of the foreign stators. We found 193 unique mutations in 87 genes, which indicated that about 2% of the 4140 genes *E. coli* harbours (KEGG database) was mutated during the evolution of (improved) motility in semi-solid agar. 54% of these mutations (104/193) were identified in 30 genes (34% of the mutated genes) we expected to influence motility (BFM, chemotaxis, PMF and peptidoglycan).

The second question focussed on the ongoing debate about the relative importance in evolution of changes in protein regulation as opposed to functional changes [41–43]. In this study, we found the majority of mutations in coding regions to be located in structural genes (155), of which 17 mutations were synonymous mutations. As synonymous mutations are more likely to affect protein regulation instead of protein function [10], the number of mutations that might affect protein function is 138 (155-17=138). These results indicated slightly more mutations resulting in functional changes (138 mutations) as opposed to mutations resulting in regulatory changes (107 mutations, of which 33 mutations were located in regulatory genes), assuming that all mutations in non-coding regions (57) have a role in regulation.

The third question focussed on the level of clustering of mutations in the structural genes and around the foreign stators in terms of physical interactions. When we investigated the number of unique mutations in each of the focal genes, we found most mutations in MotB homologues. In order from most mutations to least mutations, we found MotB (21 mutations), MotA (18 mutations), FliG (14 mutations), FliM (11 mutations), CheB (6 mutations), FliC (4 mutations), CheW (2 mutations), FliH/FliG/CheA/Tsr (1 mutation). We also examined the number of mutations per gene at the level of sub-classes in the structural genes. Here, we focussed on the stator complex, the BFM without stator complex, the chemotaxis signalling pathway, proteins that affected the PMF and proteins that affected the peptidoglycan layer. Considering the number of mutations that occurred in components/systems involved in different cell processes, the results showed that most mutations per gene were found in the stator complex (19.5), followed by the other BFM components (6.2), the chemotaxis signalling pathway (2.5), peptidoglycan-affecting genes (2) and PMF-affecting genes (1.1). Together, this indicated that mutations were most densely clustered in the structural vicinity of the stator proteins, especially affecting BFM proteins.

3.3.3 Between-strain comparison of genome-wide mutations

In the previous section we discussed the genome-wide pattern of unique mutations without making a distinction between different cBFM strains. Because different foreign stators might have led to different selectively-accessible, mutational trajectories, we next compared the genome-wide pattern of mutations between different cBFM strains. Similar to the results above, we started with the distinction between unique mutations in non-coding and coding regions and then zoom in to the single-gene level.

3.3.3.1 Mutations in coding versus non-coding regions

All six cBFM strains that evolved (improved) motility in semi-solid agar harboured more unique mutations in coding regions than in non-coding regions. Strains 08 and 29 contained a relatively high number of mutations in non-coding regions compared with the other strains (Figure 3.2). Because some cBFM strains generated more lineages (by branching) than others, comparing only the number of unique mutations between strains might not be the most meaningful analysis. In order to make a more meaningful comparison, we normalized the number of unique mutations per cBFM strain by dividing it by the number of endpoints that were sequenced. The results showed that the number of unique mutations per endpoint ranged from 1.22 in strain 08 to 3.83 in strain 13 (Figure 3.2) and indicated between-strain differences. Although this value does not take into account the number of flares that evolved and the length of evolutionary lineages leading to the endpoints (i.e. between many short lineages or few long lineages), it is indicative of the amount of genetic change that occurred during the selection experiment. In all strains we found many more SNPs than insertions or deletions (InDels), a result not surprising giving the low number of indels we found in general (Figure 3.1 top left). Only strain 08 lacked InDels among the mutations that were discovered.

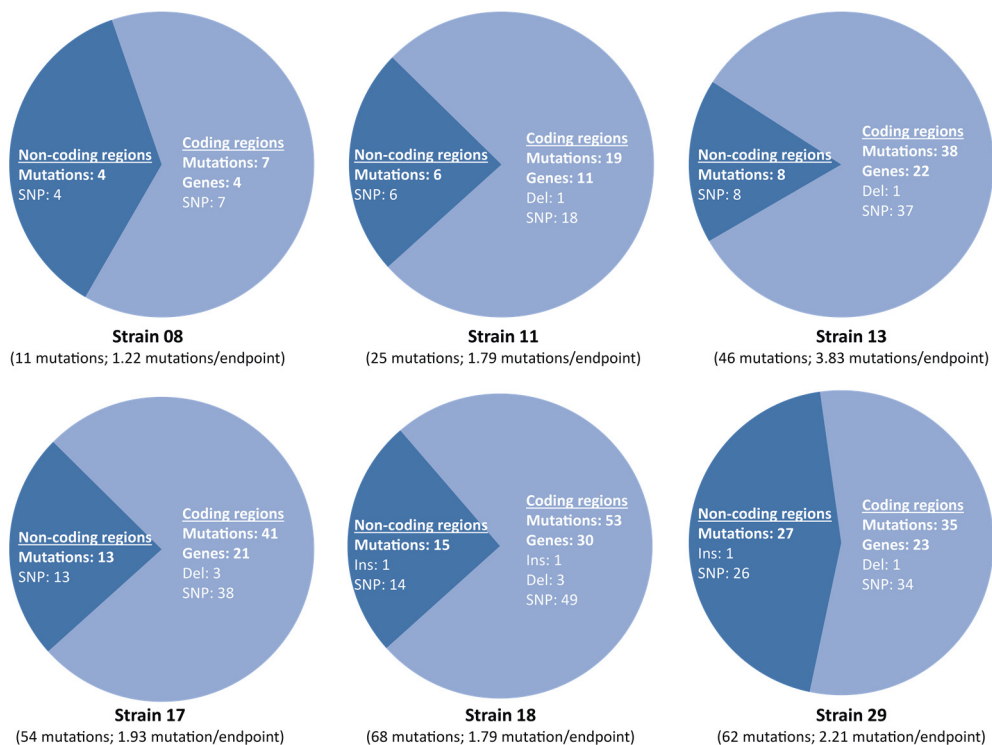


Figure 3.2. Distribution of unique genome-wide mutations over coding and non-coding regions in cBFM strains that evolved (improved) motility. The size of pie-chart sections represents the number of mutations.

Taken together, these results show that, although the overall pattern of mutations was similar between cBFM strains, they differed in terms of the number of unique mutations that were acquired during adaptive evolution on semi-solid agar plates. This is consistent with the picture that emerged from the evolutionary dynamics and phenotypic analysis presented in Chapter 2, which revealed that integration of different foreign stator complexes involved different mutational routes.

3.3.3.2 Mutations in regulatory versus structural genes

Next, we focused on the distribution of mutations between regulatory and structural genes. We already described that the overall pattern of mutations showed more mutations in structural genes (Figure 3.1), but does this apply to all cBFM strains?

As indicated in Figure 3.3, different cBFM strains showed different distributions of mutations between regulatory and structural genes (for specific genes, see Table 3.3 to Table 3.8 in the supplementary information). While strain 08 lacked mutations in regulatory genes, strain 11 gained more mutations in regulatory genes than in structural genes. These two strains represented the two most distinct strains. The remaining four strains all contained more structural genes that were mutated, although the ratio of the two types of genes varied greatly between strains. In three strains (13, 18 and 29), we furthermore identified mutations in genes of which the function is not annotated yet.

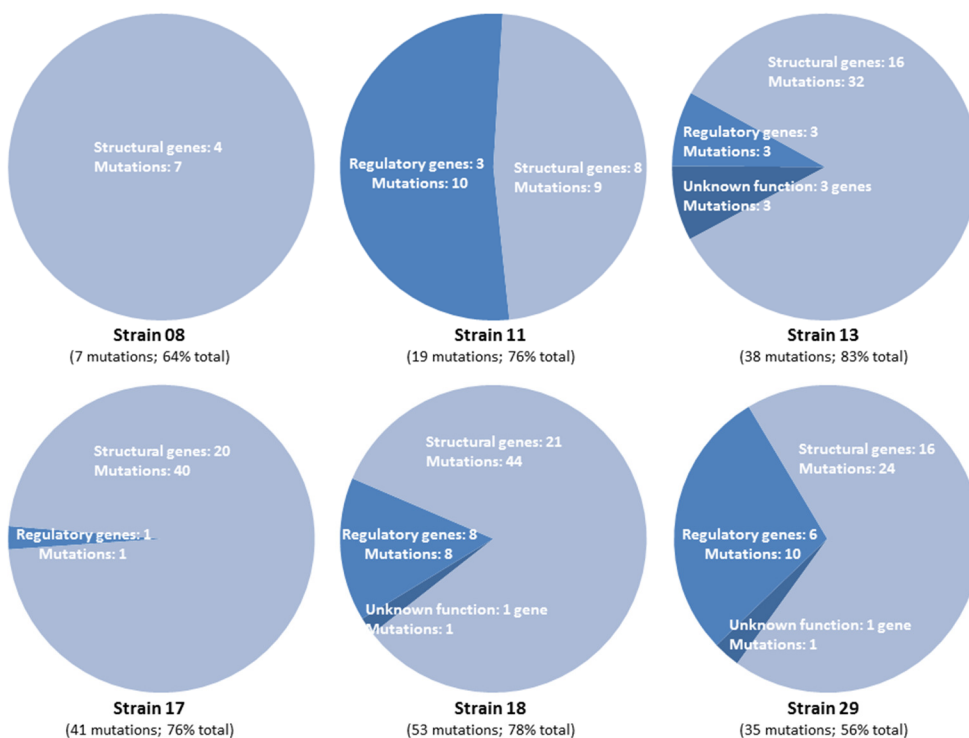


Figure 3.3. Distribution of unique mutations over regulatory and structural genes in all cBFM strains that evolved (improved) motility. The size of pie-chart sections represents the number of mutations. The percentage indicates the fraction of the total number of mutations for each strain.

These findings indicated that the different mutational routes of strains with foreign stators that evolved increased motility, differed greatly in terms of the relative importance of regulatory versus functional changes.

3.3.3.3 Mutations in structural genes

Next we focussed on the distribution of mutations in the class of structural genes, with a distinction between BFM, chemotaxis, PMF, peptidoglycan and other classes of genes.

In most cBFM strains, mutations in the BFM genes constituted at least one third of the mutations in structural genes (Figure 3.4; for specific genes, see Table 3.3 to Table 3.8 in the supplementary information). Strain 11 was an exception to this, with less than 25% of the mutations being located in BFM genes.

Chemotaxis genes were not mutated in strains 08 and 11, with only strain 13 harbouring more than 2 mutations in these genes. Although tuning chemotaxis can be important for motility of populations of cells in semi-solid agar (e.g. by fine tuning the relation between single cell swimming speed and the ability to follow an attractant gradient), the lack of mutations in two strains might indicate that changes in these genes were of less general importance than changes in BFM genes. This suggests that the different foreign stators induced different kinds of interaction conflicts.

Similar to tuning chemotaxis, optimising the PMF might be a way of increasing the motility of populations of cells in semi-solid agar. As proton translocation is essential for torque generation and thus motility [21], changing the PMF, which influences proton translocation, might affect motility. The results showed that PMF genes were mutated in all strains. However, in most cases different PMF genes were affected in different cBFM strains (*atpA* is an exception), which included genes that were, for instance, involved in intracellular electron transport, ATP synthesis or (protein) transport across the membrane (for specific genes, see Table 3.3 to Table 3.8 in the supplementary information). These results highlight the importance of the PMF genes in the evolution of motility and indicated there might have been multiple genetic routes toward improved motility by modification of the PMF.

Mutations in peptidoglycan genes might have an effect on motility of populations of cells, if these, for instance, improve the positioning of the foreign stator around the rotor. Since MotB interacts with the peptidoglycan layer, this scenario is not unlikely. Most strains, strains 08 and 11 were the exception, were found to harbour mutations in peptidoglycan genes, similar to what we found for the chemotaxis genes (for specific genes, see Table 3.3 to Table 3.8 in the supplementary information). Furthermore, one peptidoglycan gene was mutated in two different cBFM strains (*dacA*). This result indicated that mutations in peptidoglycan genes might be important in the evolution of (improved) motility, but also that it might depend on the identity of the foreign stator.

In all six strains we also discovered mutations in other structural genes, which included genes that affected antibiotic resistance, amino acid synthesis or DNA recombination (for specific genes, see Table 3.3 to Table 3.8 in the supplementary information). In some instances, we found the same gene to be affected in different cBFM strains (e.g. *yceG* and *sanA*). However, based on the functional descriptions available for these genes, there was no easy way to define a functional relationship with

motility. Therefore, we left these genes out of further analysis. Nonetheless, these observations underline the emerging picture that different stators imposed different constraints on the range of accessible adaptive mutations.

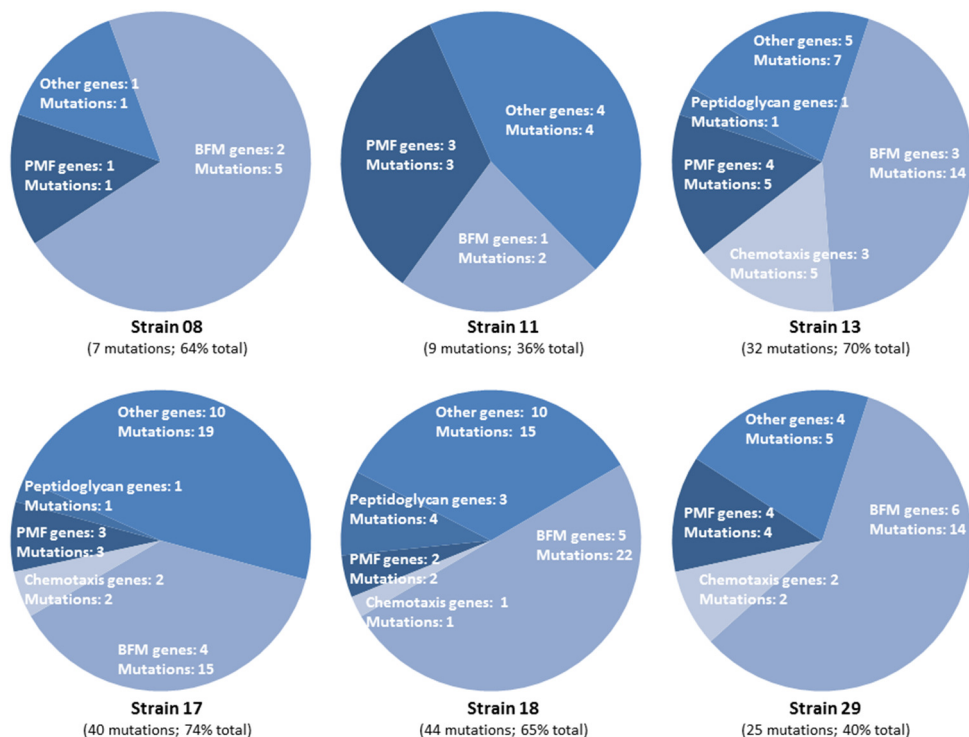


Figure 3.4. Distribution of unique mutations over structural genes in all cBFM strains that evolved (improved) motility. The size of pie-chart sections represents the number of mutations. The percentage indicates the fraction of the total number of mutations for each strain.

To summarize, the majority of mutations in structural genes was located in BFM, chemotaxis, PMF and peptidoglycan genes, which implied that mutations were clustered in all six strains. However, the distribution of mutations between these classes indicated strain-specific differences. In the next section we will investigate the pattern of mutations within the BFM genes in more detail.

3.3.3.4 Mutations in BFM genes

Our observations revealed that seven different BFM genes were mutated in total in all six cBFM strains, with either one or both of the stator genes being mutated in all cBFM strains (

Figure 3.5; for specific genes, see Table 3.3 to Table 3.8 in the supplementary information). The percentage of the total number of unique mutations that had occurred in each strain differed considerably between cBFM strains, ranging from 8% for strain 11 to 45% for strain 08, with the majority of strains having a percentage of around 30%. We also found a large variation between cBFM strains in the number of different BFM genes that contained mutations, ranging from one

mutated gene in strain 11 to six mutated genes in strain 29. Furthermore, we discovered that strains 13, 18 and 29 harboured many mutations in other BFM genes besides mutations in *motA*, *motB* and *fliG*, in contrast to strains 08, 11 and 17. Previously (Chapter 2), we described data that suggested that strains 13, 18 and 29 did not have access to big-benefit mutations early on during evolution. Big-benefit mutations were defined as mutations that conferred an increase in relative population swimming speed of >50% of the difference between the highest speed obtained by an endpoint in that strain and the initial speed. The lack of big-benefit mutations may have allowed for mutations (of smaller effect) in other BFM genes, besides *motAB* and *fliG*, to have an opportunity to emerge and accumulate. Consequently, it appears to be the case that big-benefit mutations especially occur in the stator genes and *fliG*, similar to what was found before for compensatory mutations. In [33], the authors showed a weak negative correlation between the effectiveness of compensatory mutations and the distance between the site of compensatory and deleterious mutations in terms of amino acid sequence. The shorter the distance, the larger the effect. In our case this seems to be the case for structural vicinity also.

Apart from identical genes that were mutated in different strains, in a few cases identical amino acids were mutated. For instance mutation FliG Q280R, a mutation found in strains 08, 17 and 18, and being located at the MotA-FliG interaction site (see also section 3.3.5). Also FliM contained identical mutations that were found in different strains, i.e. E49K in strains 18 and 29, A161V in strains 13 and 18 and N201H in strains 13 and 29.

Taken together, above results indicated both differences and similarities between different cBFM strains, although identical mutations might still have different effects due to differences in genetic background (epistatic interactions). For instance, residue 280 in FliG interacts with residues of MotA [12,24–26]. But structural differences between different MotA proteins (e.g. MotA from strain 08 and 17) might cause residue 280 of FliG to have a different interaction and thus, potentially, a different effect of mutation.

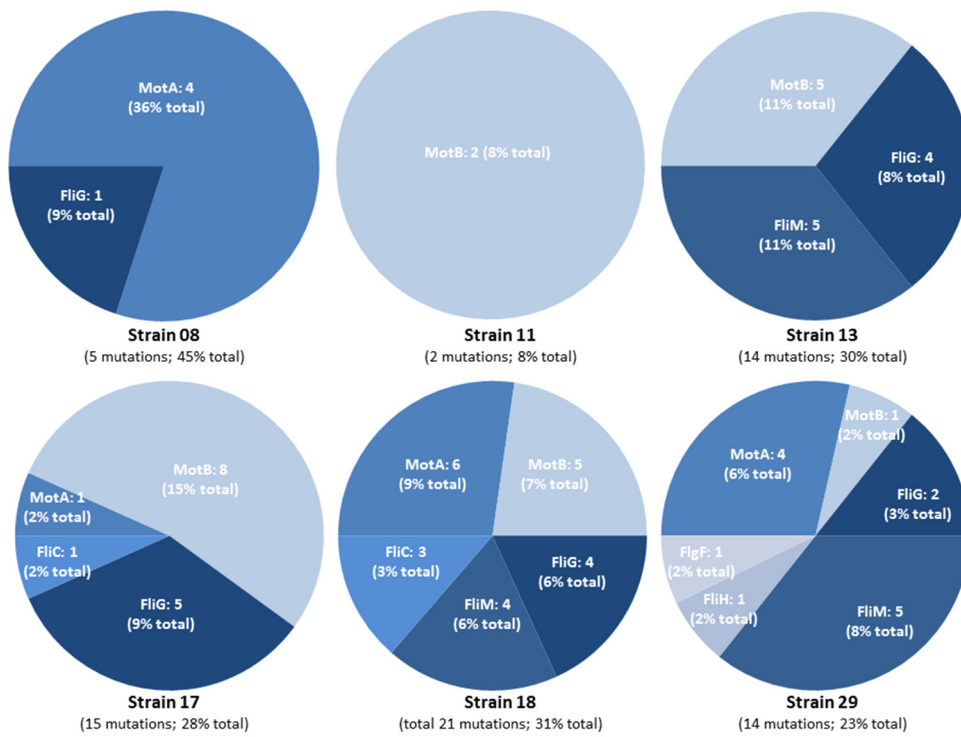


Figure 3.5. Distribution of unique mutations over BFM genes in all cBFM strains that evolved (improved) motility. The size of pie-chart sections represents the number of mutations. The percentage indicates the fraction of the total number of mutations for each strain.

3.3.3.5 Mutations in chemotaxis signalling pathway genes

Only strains 08 and 11 did not harbour mutations in the chemotaxis genes (Figure 3.6; for specific genes, see Table 3.3 to Table 3.8 in the supplementary information). *CheB* was found to be mutated in four strains, although all mutations were different. Strains 17 and 29 were mutated in *cheW* in addition to *cheB* and strain 13 harboured additional mutations in *cheA* and *tsr*. Strain 18 did not gain additional mutations besides mutations in *cheB*, complementing the between-strain diversity in mutations in chemotaxis genes.

Mutations in chemotaxis genes were only discovered in strains that evolved more slowly (i.e. needed more time to reach high motility) and showed longer evolutionary lineages on average. This indicates that mutations in chemotaxis genes may only be involved in the integration of the foreign stators when no big-benefit mutations early in evolution were available. This is similar to the case discussed above, where mutations in BFM genes outside of the stator genes and *fliG* were seen when no big benefit mutations were accessible early in the trajectory. Although mutations in chemotaxis genes were also present in strain 17, this strain contained two long evolutionary lineages that could harbour these mutations (see section 3.3.4.1).

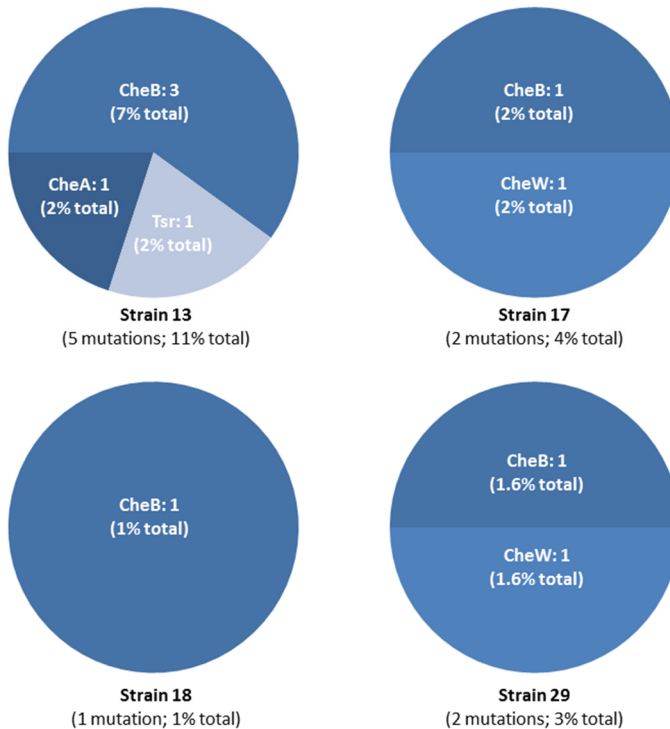


Figure 3.6. Distribution of unique mutations over chemotaxis genes in four cBFM strains that evolved (improved) motility. The size of pie-chart sections represents the number of mutations. The percentage indicates the fraction of the total number of mutations for each strain.

3.3.4 Within-strain comparison of genome-wide pattern of mutations

In the previous sections we compared the genome-wide patterns of mutations that were discovered by whole-genome re-sequencing between different cBFM strains. However, the endpoints of strains 17, 18 and 29 were distributed over two or more pools that were sequenced. Distribution of the replicate-line endpoints was based on differences in the length of evolutionary lineages (discussed in Chapter 2) and therefore we expected to find genetic differences between pools.

3.3.4.1 Strain 17

During the selection experiment, strain 17 evolved 28 endpoints, which were distributed over two pools. As was shown previously (Chapter 2), half of the endpoints evolved from two strains isolated from primary flares (17.1d and 17.1g). These endpoints were combined in pool 1, while the remaining endpoints were combined in pool 2 (Figure 3.7). As such, the average length of evolutionary lineages was 3.75 flares for pool 1 and 1.29 flares for pool 2. Operating under the assumption that longer lineages harbour more mutations than short lineages and according to our big-benefit mutation hypothesis, this would lead to different mutations in both pools of strain 17. In general, we expected to find more mutations in pool 1 than in pool 2, because of the longer

evolutionary lineages. Furthermore, especially in pool 1 we expected to find mutations in other genes besides the stator genes and *fliG*, for instance the above-mentioned mutations in the chemotaxis genes.

As expected, we found that the number of mutations was different between both pools (Figure 3.7A). While 40 mutations were found in pool 1 with the longer evolutionary lineages (2.86 mutations per endpoint), 29 mutations were found in pool 2 with the shorter lineages (2.07 mutations per endpoint). The distribution of mutations over non-coding and coding regions in the two pools was similar, as was the distribution of mutations between regulatory and structural genes (Figure 3.7B). Two of these mutations, one in a regulatory gene and one in a structural gene, were found in both pools (Table 3.4 in supplementary information). More differences between pool 1 and 2 were discovered, when we compared the different classes of structural genes (Figure 3.7C) as well as the *BFM* (Figure 3.7D) and chemotaxis genes (Figure 3.7E). For instance, mutations in peptidoglycan genes, chemotaxis genes and *fliC* were found in pool 1, but were absent in pool 2. Furthermore, most discovered mutations in *motB* and *fliG* were different between both pools (Table 3.4 in supplementary information). Only one mutation in *motB* was found in both pools.

Above results showed mutational differences between replicate selection lines of strain 17, which indicated that the same foreign stator could be integrated via different mutational routes. Furthermore, the results presented additional evidence of our big-benefit mutation hypothesis. In pool 1, which contained genotypes that lacked big-benefit mutations early in evolution, we found mutations in other BFM genes outside of the stator genes and *fliG* and mutations in chemotaxis genes. These mutations were absent in pool 2.

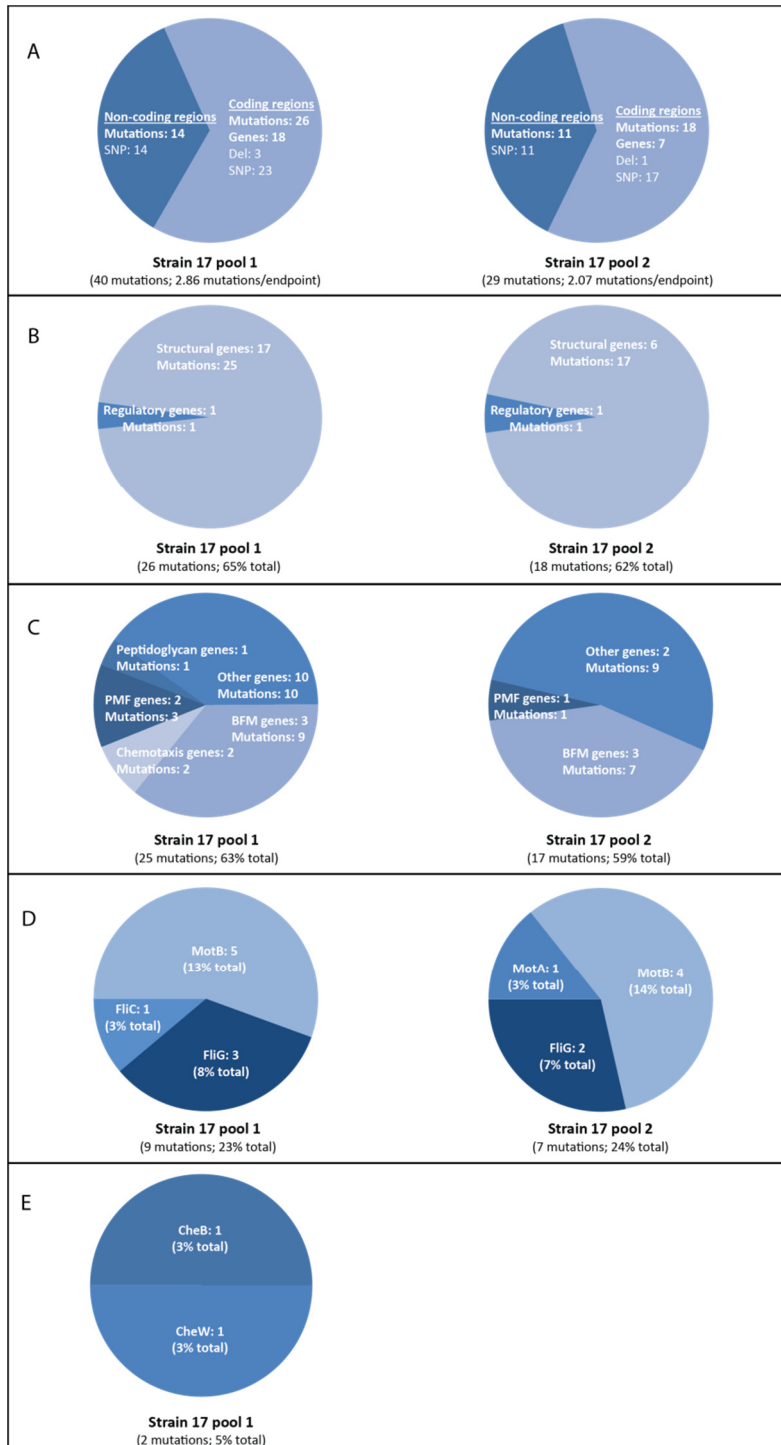


Figure 3.7. Distribution of mutations in two pools of strain 17 that contained end-points of long versus short lineages. A. Non-coding vs coding mutations. B. Mutations in regulatory genes vs structural genes. C. Several classes of structural

genes. D. BFM genes. E. Chemotaxis genes. The size of pie-chart sections represents the number of mutations. The percentage indicates the fraction of the total number of mutations for each strain.

3.3.4.2 Strain 18

One of the cBFM strains that lacked big-benefit mutations early in evolution in all lineages was strain 18. This strain evolved 38 endpoints that we divided in 2 pools of 12 and 1 pool of 14 endpoints. As was shown previously (Chapter 2), the diversity in lengths of evolutionary lineages that evolved in strain 18 was great. By using this diversity when composing the pools, this resulted in between-pool differences in average lengths of evolutionary lineages. The average lengths were 2.75 flares for pool 1 (12 endpoints that evolved from genotypes isolated from primary flares 1b, 1c, 1e, 1h, 1j, 1l), 4.00 flares for pool 2 (14 endpoints that evolved from genotypes isolated from primary flares 1a, 1d, 1i, 1n) and 2.33 flares for pool 3 (12 endpoints that evolved from genotypes isolated from primary flares 1f, 1g, 1k, 1m).

Although the average lineage length of pool 1 was lower than that of pool 2, it contained a higher number of mutations per endpoint (Figure 3.8A; 3.58 and 3, respectively), with a similar distribution of mutations over non-coding and coding regions. Pool 3, with the shortest average lineage length, contained the lowest number of mutations per endpoint. However, in contrast to pools 1 and 2, it contained relatively few mutations in coding regions, while the number of mutations in non-coding regions was similar to the other pools (Figure 3.8A). These results showed that the mutational trajectories were less intuitive than expected before: shorter lineages not always harboured less mutations than longer lineages.

At the level of regulatory and structural genes (Figure 3.8B) no large differences between pools were found, but zooming in on the structural genes only (Figure 3.8C) unveiled additional differences. Pool 3 only harboured a few mutations in BFM genes and loci of the *other genes* category, while pool 2 also harboured mutations in PMF and peptidoglycan genes. The highest number of different, mutated structural genes were discovered in pool 1. This pool contained mutations in the same categories as pool 2, but additionally contained a mutation in chemotaxis genes. Zooming in further on the BFM genes (Figure 3.8D) showed that both stator genes were mutated in all three pools. Of the mutations in BFM genes, four mutations were present in more than one pool (Table 3.5 in supplementary information). Although mutations were found in *fliG*, *fliM* and *fliC* in both pool 1 and 2, only 1 of the 10 mutations identified was present in both pools. This mutation was A161V in *fliM*. Finally, a single mutation was discovered in the chemotaxis genes (Figure 3.8E; pool 1).

Together, these results indicated that the mutational possibilities to improved motility in strain 18 were diverse, similar to strain 17. In Chapter 2 we identified for strain 18 a positive correlation between the length of an evolutionary lineage and relative population swimming speed. In other words, in the absence of big-benefit mutations early in evolution, we found, on average, that a longer evolutionary lineage corresponded to a higher motility. Together with the results discussed above, we showed that not all mutations in *motAB* provided big-benefits, as pool 3 had a low population motility on average (Chapter 2). Furthermore, we again showed that mutations in other genes besides the stator genes were involved in reaching higher motility in the absence of big-benefit mutations early in evolution (pool 1 and 2). With this information, we can thus define our big-benefit hypothesis more accurately. Big-benefit mutations especially occur in the stator genes and *fliG*,

although not all mutations in the stator genes result in big benefits. The absence of big-benefit mutations early in evolution provides the opportunity for motility-increasing mutations in other genes to emerge. In addition, it raises the possibility that early mutations in *MotAB* prevent subsequent big-benefit mutations in *MotAB*, as follows from the observation of mutations in these genes in pools containing endpoints with low swimming speeds.

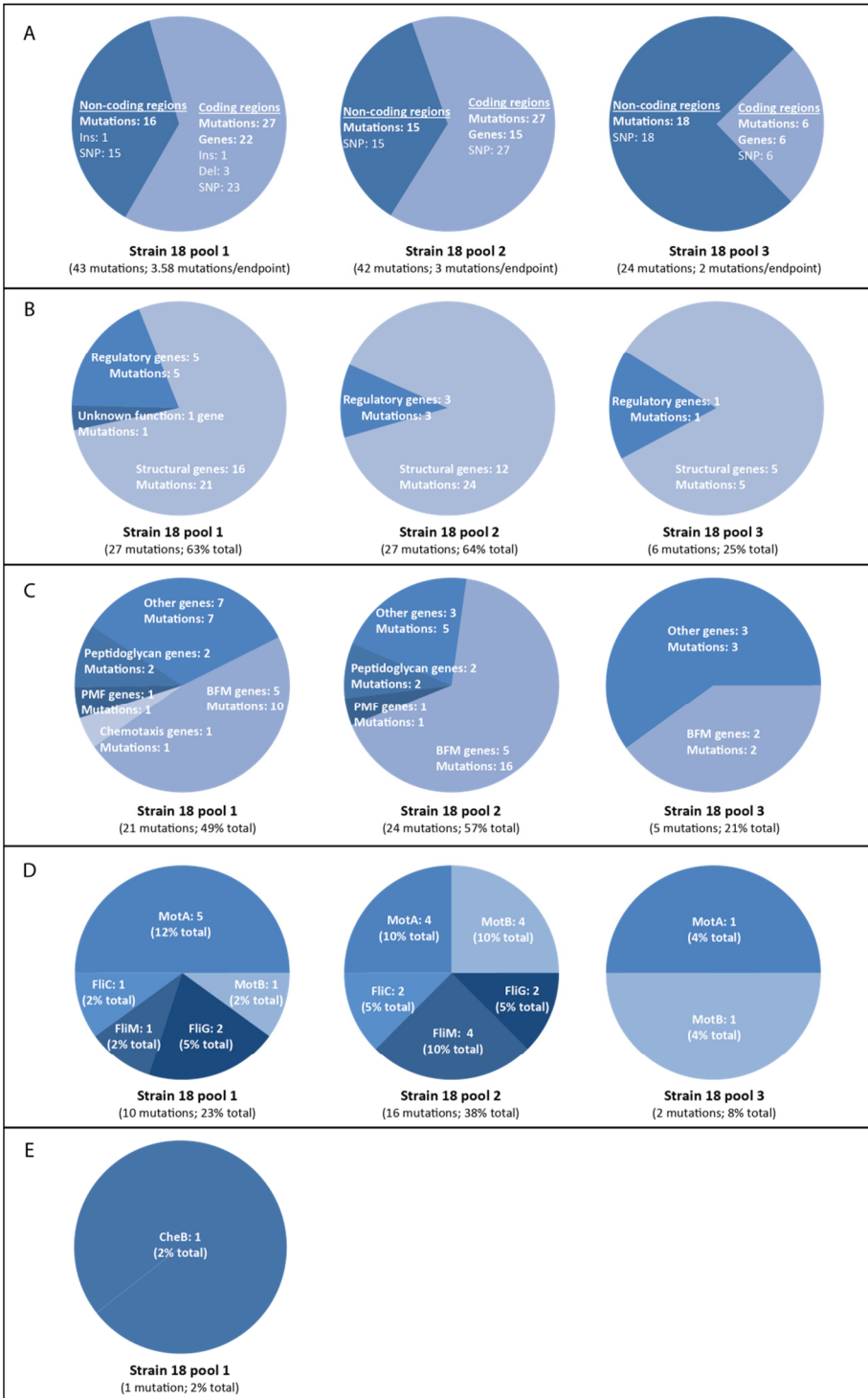


Figure 3.8. Distribution of mutations in three pools of strain 18 that contained end-points of long versus short lineages. A. Non-coding vs coding mutations. B. Mutations in regulatory genes vs structural genes. C. Several classes of structural genes. D. BFM genes. E. Chemotaxis genes. The size of pie-chart sections represents the number of mutations. The percentage indicates the fraction of the total number of mutations for each strain.

3.3.4.3 Strain 29

Similar to strain 18, strain 29 also lacked big-benefit mutations in all lineages early in evolution. Both pools of strain 29 consisted of 14 endpoints. The average length of evolutionary lineages in pool 1 was 2.87 flares (endpoints that evolved from genotypes isolated from primary flares 1a, 1d, 1g, 1j, 1l, 1m) and the average length of pool 2 was 4.00 flares (endpoints that evolved from genotypes isolated from primary flares 1b, 1c, 1e, 1f, 1h, 1i, 1k). Due to these differences in average length of evolutionary lineages, we expected to find mutational differences between pools with more mutations in pool 2.

In agreement with our expectation, pool 2 contained more mutations per endpoint than pool 1 (Figure 3.9A; 3.21 and 2.29, respectively). However, the distribution of mutations between coding and non-coding regions showed no large differences between pools. In contrast to other cBFM strains that evolved (improved) motility, strain 29 harboured mutations in the SD sequence of *motA* and in the p15A origin of replication of the plasmid that contained the stator genes. These mutations were present in both pools, with three unique mutations in pool 2. Pool 2 also harboured more mutations that affected regulatory genes compared with pool 1 (Figure 3.9B), which indicated that these mutations only seemed to occur in the longer evolutionary lineages. Zooming in on the structural genes (Figure 3.9C), we found chemotaxis genes only to be mutated in pool 2, while mutations in BFM genes formed the majority of the mutations in structural genes in both pools. Many identical mutations in the stator genes were found (Figure 3.9D and Table 3.8 in supplementary information), which was not the case for other BFM genes.

Similar to strains 17 and 18, mutational differences between replicate selection lines were found for strain 29. Furthermore, the data is in agreement with our big-benefit mutation hypothesis: in absence of such mutations early in evolution, mutations outside of the stator genes and *fliG* have an opportunity to emerge.

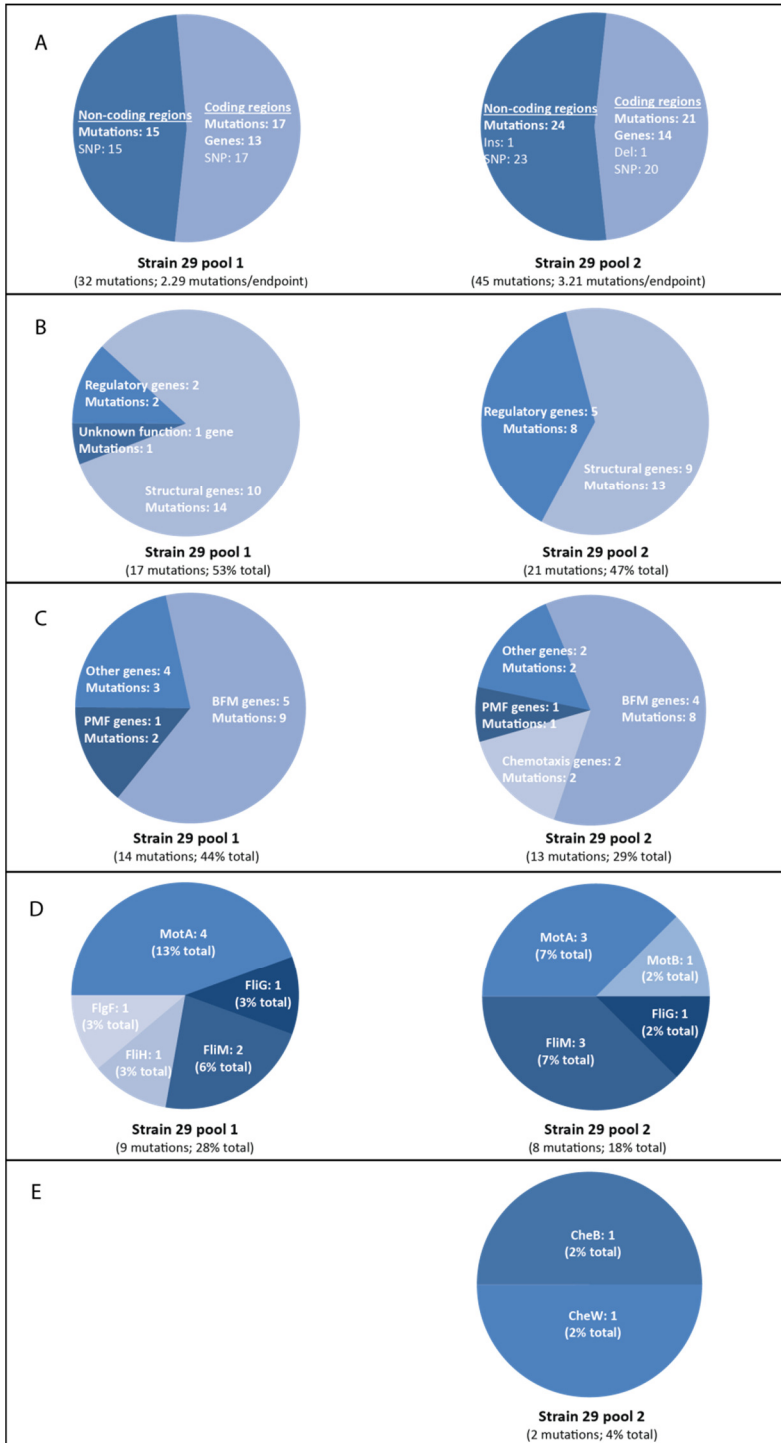


Figure 3.9. Distribution of mutations in two pools of strain 29 that contained end-points of long versus short lineages. A. Non-coding vs coding mutations. B. Mutations in regulatory genes vs structural genes. C. Several classes of structural

genes. D. BFM genes. E. Chemotaxis genes. The size of pie-chart sections represents the number of mutations. The percentage indicates the fraction of the total number of mutations for each strain.

3.3.5 Between-strain comparison of mutations in the stator complex and mutations at the MotA-FliG interaction site

In section 3.3.3, we compared the discovered mutations at the level of genes in order to examine strain-specific differences. In this section, we zoom in and compare different cBFM strains by examining the positions of individual mutations in the stator complex and in FliG at the site where it interacts with MotA. We focused on the stator complex, because this was the component that was foreign to the *E. coli* cell and also what set apart the different cBFM strains. The nature of the incompatibilities of these different stators may differ and as a result different mutations in these components may be required for their integration. Because of this, one may expect to see strain-specific mutations. We also focused on FliG at the site where it interacts with MotA, because this site is essential for flagellar rotation and torque generation [12,24–26] and important for proper positioning of the stator around the rotor [50,51]. Although a diverse set of other mutations in FliG was found, all more to the N-terminus of the FliG, we did not take this into consideration at this point. The structures shown in Figure 3.10 were obtained by predicting the transmembrane (TM) helices in MotA and MotB through the online TMHMM Server v.2.0 [52]. For clarity, the numbers of amino acid positions at the start and end of TM helices were omitted. Alignment of the foreign stator sequences to the replaced stator sequences of *E. coli* resulted in the predicted positions of mutated amino acids. As gaps were involved in all alignments, the actual positions are likely to be different. However, the regions in which mutations occurred (e.g. TM helix or linker-domain between TM helices) are likely to be accurate.

The distribution of mutations within the stator proteins differed between cBFM strains (Figure 3.10). Although some strains showed identical mutations (e.g. FliG Q280R) and overlap in mutated regions (e.g. the plug domain of MotB), they all harboured also unique mutations. Thus, evolution of (improved) motility at the level of MotAB and FliG involved different subsets of mutations in each strain. The MotA mutations in strains 08 and 18 were located along the cytoplasmic loop, including the MotA-FliG interaction site, and the C-terminal domain. Additionally, strain 18 also harboured mutations in MotB. These mutations were all located C-terminally of the TM helix of MotB, including mutations in the plug domain and peptidoglycan binding domain (PBD). The plug domain is involved in the regulation of the flow of ions through the stator complex, which it prevents when that stator is not bound to the BFM, and might also affect the interaction of the PBD with the peptidoglycan layer upon docking to a BFM [53]. The PBD might affect the positioning of the stator around the rotor [35]. Mutations in these two domains were also found in strain 11, although at different positions. Similar to strains 11 and 18, strain 17 harboured mutations in the plug domain of MotB. Additionally, unique mutations not discovered in other strains were found in strain 17 in the TM helix of MotB and in the third TM helix of MotA.

The only strain that harboured mutations at the N-terminus of MotA was strain 29, of which three mutations were synonymous mutations. An additional mutation was found in the plug domain of

MotB, at a position that was similar by alignment to the A52T mutation in strain 17. Strain 13 harboured five mutations in the stator proteins, all of which were located to the PBD of MotB.

Besides mutations in MotA at the MotA-FliG interaction site, we also discovered mutations at this site in FliG. One mutation, Q280R, was found in three cBFM strains (strains 08, 17 and 18), which indicated parallel evolution due to similar conflicting interactions of these different foreign stator. Two other mutations were found in close vicinity of the interaction site and were discovered in strains 13 and 17.

Strains 17, 18 and 29 allowed us to compare mutations between replicate selection lines, as we had more than one pool for each strain. In all strains, between-pool differences were found in the mutations that affected *motAB* and *fliG* at the site where it interacts with MotA. Furthermore, strain 17 allowed us to identify possible candidates for big-benefit mutations, which were located in both stator genes as well as *fliG* (Table 3.4 in supplementary information). We suggest the following mutations as candidates for big-benefit mutations: I168M in MotA, S39R and F53V/L in MotB, Q280R in FliG. Furthermore, in Chapter 5 we will see that in both pools of strain 17 large indels were present, some of which (the deletions in MotB) conferred a large fitness benefit.

In summary, besides finding mutations that were different between cBFM strains, indicating strain-specific constraints as expected, we also discovered mutations at sites that were either identical (in FliG) or aligned to the same amino acid in *E. coli*. Furthermore, we found that different cBFM strains often harboured mutations in the same regions of the stator proteins, possibly indicating similarities in the mechanisms of integration.

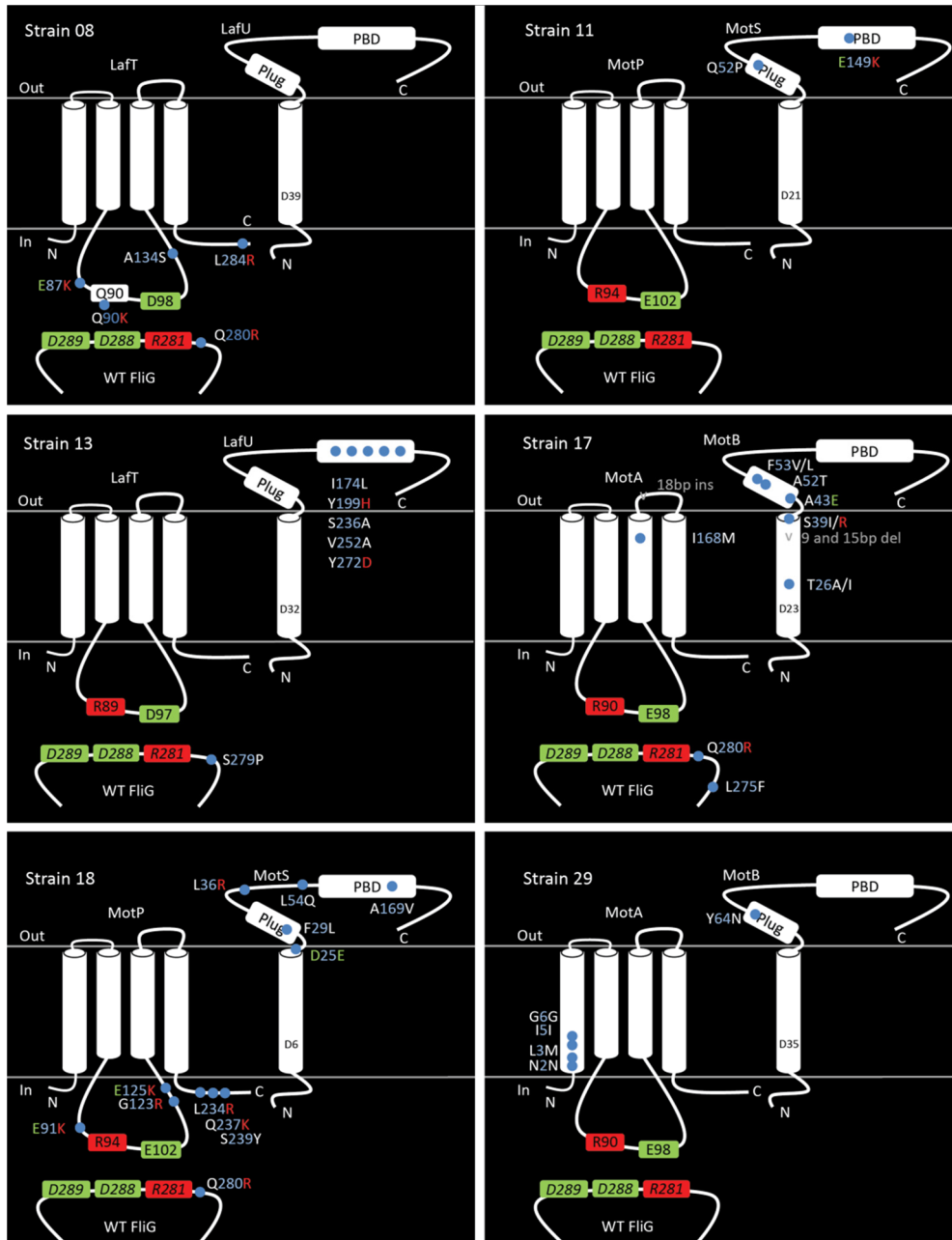


Figure 3.10. Mutations in the stator proteins and in the MotA interaction site of FlIG. Transmembrane domains are represented by cylinders. Residues in red are positively charged. Residues in green are negatively charged. Residues in white are uncharged. Strain 17 contains insertions (ins) and deletions (del), which were not identified by NGS (see Chapter 5). Horizontal lines indicate the cell membrane that separates the cytoplasm (in) from the periplasmic space (out). PBD: peptidoglycan binding domain. Plug: plug domain. Positions of mutations, plug and PBD were determined by alignment to *E. coli* stators. See text for details.

3.4 Discussion

Experimental data about the underlying molecular mechanisms that facilitate the evolution of protein complexes by compositional evolution is scarce. In this study we replaced the stator genes of *E. coli* with foreign stator genes from different species across the bacterial domain and investigated the presence and location of mutations that occurred during adaptive evolution on semi-solid agar plates that favoured increased motility. In order to do this, we performed a selection experiment (Chapter 2) followed by whole-genome re-sequencing, which included the plasmids that harboured the foreign stator genes. We were interested not only in within-strain diversity in mutations, but also in between-strain differences. Furthermore, we were interested in the clustering of mutations. Previous studies that investigated compensatory mutations, which might have similar effects as the mutations in our study, showed that these mutations tend to be clustered around the site of the initial deleterious mutation [32] and that those mutations tend to be more effective [33]. Furthermore, studies of deleterious mutations in the stator and rotor of the BFM, showed that compensatory mutations could be located in both stator and rotor [34–37,54,55].

Without reconstructing mutations, it is not possible to exactly indicate the role of these mutations in an observed phenotypic change. Hence, we cannot distinguish between mutations that increased motility by directly improving integration of the foreign stator in the cell and mutations that increased motility by a different mechanism that had no direct functional relation to the foreign component (and thus led to indirect improvement of integration of the foreign stator in the cell). Furthermore, some mutations might have hitchhiked with a beneficial mutation and might be neutral or even have a negative effect themselves. Also, it is important to point out that our selection regime, in addition to favouring more motile genotypes, can have selected for changes in other phenotypes. Nonetheless, given our selection regime - that is, selecting for flares or cells that end up in the population front on semi-solid agar - and the fact that our founding cBFM strains were not or very poorly motile, we can assume that many of the mutations we identified, and in particular those in the focal genes, were involved in adaptive evolution of motility in semi-solid agar. A second point that should be noted is that the adaptive consequences of changes in structural genes can involve any combination of changes in the functioning of the encoded protein and changes in its expression level.

3.4.1 Overall pattern of mutations

The average number of mutations per endpoint of an evolutionary lineage is 2.13. Even when we consider the lower boundary of the 95% confidence interval of found mutations (77.3%), the average number of mutations per endpoint (1.65) is consistent with the conclusion that, based on the patterns of phenotypic evolution, we have captured multi-step evolutionary trajectories.

3.4.1.1 Structural versus regulatory evolution

There has been considerable discussion in the literature, regarding the relative importance of mutations that affect gene function vs gene regulation in the process of adaptive evolution [41–44]. We found that the majority of mutations was located in structural genes (155 mutations) as opposed to regulatory genes (33 mutations; Figure 3.1). However, some mutations in structural genes were synonymous mutations. This implied that, if they conferred a selective advantage, this would have

involved a regulatory effect [10]. Also, mutations in non-coding regions could affect regulatory processes. If we consider both of these two classes of mutations to affect regulation, then the number of mutations with functional or regulatory effects were quite similar (138 and 107 respectively). In spite of the interpretation of the effects of these latter types of mutations, the findings show that in our model system, adaptation involved both considerable roles for regulatory and structural mutations. As mentioned above, we cannot distinguish between mutations that increased motility by directly improving integration of the foreign stator in the cell and mutations that increased motility by a different mechanism that had no relation to the foreign component whatsoever. As a result, it is not yet possible to conclude whether foreign component integration involved both structural and regulatory mutations. This would require genetic reconstruction of the mutations in the wild-type genetic background and analysis of the resulting fitness effects. Nonetheless, the observation that the ratio of mutations in structural vs. regulatory genes varied considerably between cBFM strains with different components (e.g. between strain 17 and 18), suggests that at least some of the regulatory mutations interacted epistatically with the foreign component. This suggests that functional incompatibility may at least in part be resolved by regulatory changes.

3.4.1.2 Distribution of mutations over components and cell systems

Our results showed that the mutations discovered by NGS were clustered, which is in agreement with literature about compensatory mutations [32]. Most mutations were found in the foreign stator genes (*motA* and *motB*) and genes of which the protein products are in close structural vicinity of the stator complex (Figure 3.1; *fliG* and *fliM*): 41.3% of all mutations in structural genes. Furthermore, a total of 67% of the mutations in structural genes were discovered in classes of genes that we expected to have a role in evolution of (improved) motility. Together, these results indicated that while mutations were more likely to affect specific genes, they were not restricted to the stator genes and genes of the interaction partners of the stator proteins. Our observations also indicate that, depending on the availability of big-benefit mutations, mutational modification of components further away from the foreign component are used to achieve integration.

Previously (Chapter 2), we predicted different mutational trajectories to (improved) motility within a single cBFM strain. This was based on phenotypic differences between endpoints, the fact that cells isolated from evolved populations showed a higher motility and the currently accepted paradigm that phenotypes that breed true in bacteria are genetically encoded [56,57]. Our sequencing results corroborated this prediction. However, also identical mutations were found among replicate adaptive lineages of a single cBFM strain, which indicated that evolution might use identical steps in different genetic backgrounds.

3.4.2 Between-strain differences in mutations

Next we investigated differences in mutations between cBFM strains. This result would shed light on whether selectively-accessible mutational trajectories depended on the identity of foreign stators. This result was expected from literature [58,59] and suggested by us previously (Chapter 2) due to differences in BFM structure between different bacteria [60]. Comparisons between cBFM strains indeed indicated that different strains showed different distributions of mutations across genes.

Furthermore, the identity of mutated genes also differed between strains. While BFM and PMF genes were mutated in all cBFM strains, this was different for chemotaxis and peptidoglycan genes. Similarly, the distribution of mutations within BFM genes differed between cBFM strains. While strains 08, 11 and 17 mostly did not harbour BFM-mutations outside of *motA*, *motB* and *fliG*, strains 13, 18 and 29 also contained many mutations in *fliM*, *fliC*, *fliH* and *flgF*. Additionally, mutations in chemotaxis and peptidoglycan genes were found in the latter strains. Previously (Chapter 2) we indicated that strains 08, 11 and 17 had shorter evolutionary lineages on average compared with strains 13, 18 and 29. These latter strains especially lacked one-step evolutionary lineages to high motility, possibly due to a lack of big-benefit mutations early in evolution. Taking these results together, we suggest that big-benefit mutations occurred more frequently in *motA*, *motB* and *fliG*, but that not all mutations in these genes provided big benefits. This suggestion is in accordance with literature on compensatory mutations [33], in which they found a weak relation between effect size of compensatory mutations and distance to a deleterious mutation. Our results extend this principle to multi-protein systems. Furthermore, our results imply that mutations in other BFM genes might only have been allowed the opportunity to emerge, when population swimming speed was low and big-benefit mutations were not available. The time it takes for a mutation to reach high frequency is inversely proportional to its relative effect, taking into account the expected waiting time for a mutation to appear, losses due to drift and the time to increase in frequency [61,62]. Therefore, big-benefit mutations have a higher chance of reaching high frequency than mutations with a smaller benefit.

Besides differences in mutations between cBFM strains also similarities were found. In a few cases, identical mutations were found in different strains. This indicated that different stators sometimes had similar conflicts. However, one should keep in mind that the phenotypic effects of identical mutations might have been different.

3.4.3 Mutations in stator proteins and at MotA-FliG interaction site

Mutations in the stator proteins and at the MotA-FliG interaction site were in general unique for each strain, despite a few identical mutations. The only strain that harboured mutations at the N-terminus of MotA was strain 29, of which three mutations were synonymous. These results implied that these mutations might be regulatory mutations instead of functional mutations, because synonymous mutations do not change the identity of the amino acid. As all three mutations are not favoured by the *E. coli* codon bias [63], expression might be expected to be influenced negatively. However, recent experimental data showed the opposite: rare codons at the N-terminus of proteins increase protein levels [11]. The fourth mutation caused a change in the size of the side chain, but not charge.

Six mutations were located in the cytoplasmic loop of MotA, three in strain 08 and three in strain 18. Two of these mutations, E87K in strain 08 and E91K in strain 18, might be similar in their effects. As both mutations cause a charge reversal from negative to positive, this might result in improved interaction with the negatively charged amino acids of FliG at positions 288 and 289. These mutations might thus result in improved rotor-stator positioning or more efficient torque generation [26,51]. Furthermore, we predict the Q90K mutation in strain 08 to have a similar effect, as this amino acid

aligns with R90 in *E. coli*. The remaining three mutations in the cytoplasmic loop of MotA were all located in between the MotA-FliG interaction site and the third TM helix and might cause structural changes. Previously, mutations in this region of MotA were found to suppress the negative effects of mutations in the C-terminal domain of MotA, which was suggested to be due to the proper alignment of the MotA-FliG interaction site [36].

Strain 17 harboured one mutation in the third TM helix of MotA. As TM helix 3 and 4 of MotA form the ion channel together with the TM helix of MotB [22,23], mutation I168M might cause a conformational change that affects the transport of ions across the membrane. Because ion translocation is essential for rotation of the BFM [21], this mutation might have a profound effect. The four mutations in the C-terminal part of MotA might cause structural changes that affect the MotA-FliG interaction site [36]. Additionally, L284R in strain 08 might also affect the Shine-Dalgarno site of MotB, thereby having a regulatory function [64].

All strains except strain 08 harboured mutations in the MotB homologue. Mutations T26A and T26I in strain 17 might affect ion translocation, as they are located in the vicinity of the ion-binding site of MotB [21], D23 in strain 17. Furthermore, strains 17 and 18 contained mutations at the C-terminus of the TM helix. The roles of these residues have not been studied previously, but it might be that these changes cause a structural change that either affects ion translocation or affects the positioning of the stator around the rotor.

All stators are foreign to the *E. coli* cell and therefore likely are not optimally compatible with different *E. coli* structures. Hence, we speculate that all mutations that affect the plug domain [53] and PBD [19,20] of MotB change the conformation of the stator complex, thereby adapting to the *E. coli* cell and improving the positioning around the rotor. The plug domain is involved in the regulation of the flow of ions through the stator complex, which it prevents when that stator is not bound to the BFM, and might also affect the interaction of the PBD with the peptidoglycan layer upon docking to a BFM [53]. The PBD might affect the positioning of the stator around the rotor [35].

The mutations in FliG at the MotA-FliG interaction site likely affect the efficiency of the rotor-stator interaction. One mutation, Q280R, was found in three different cBFM strains and reveals that different stators sometimes also had similar interaction conflicts in the host *E. coli*. This mutation causes a positive charge to appear next to the positively charged R281 [12], and might be involved in the interaction with the negatively charged amino acid from MotA opposite of R281, similar to what was discovered for the sodium-driven motor of *Vibrio alginolyticus* [65]. Additional mutations in FliG near the interaction site (S279P in strain 13 and L275F in strain 17) may cause conformational changes that improve the alignment between MotA and FliG.

3.5 Conclusion

Our results reveal that integration of foreign stators could occur along multiple mutational trajectories. This corroborates the pattern of phenotypic diversity between the end points of replicate adaptive trajectories presented in Chapter 2. If we define the complexity of an evolutionary transition in terms of the number of possible solutions by which it can be achieved from a given starting point, then this finding shows that the component incompatibilities that we sampled in this experiment were not of the highest complexity class solvable by evolution. It should be noted,

however, that different trajectories might impose different constraints on the maximum compatibility that can be achieved.

The different stators of which the integration was examined in this study diverged from a common ancestral stator in evolutionary history. Our observation that the integration of these different stators occurred along mutational trajectories that were at least in part different, suggests that the homologous components diversified functionally to a point where their integration required specific mutational changes. This reveals that the strength and specificity of the interactions between a stator and the BFM and other cell systems is such that it can cause diversified components to require unique changes for their integration.

The mutations that improved the functionality of the foreign stator genes were clustered: they were located mostly in the new component and its nearest interaction partners. Moreover, the mutations that conferred the highest adaptive were often located in these components. Mutations in more distant components also occurred and in some cases appeared to have fixed in some lineages owing to epistatic constraints, engendered by earlier mutations that altered the effect of the big benefit mutations observed in other lineages.

Together, these observations began to shed some light on the nature of the evolutionary independence (epistatic modularity) of the interactions between the stator complex and the BFM and associated cellular systems. The picture that is developing reveals a compatibility horizon at which foreign components can be integrated along multiple step-wise adaptive trajectories that are differentially constrained for different homologs candidate parts.

3

3.6 Acknowledgements

We thank B. Hubert for his valuable input in discussions and his hard work in designing DNA analysis software Antonie. We also thank S. van der Hoff for performing many tests in the lab.

3.7 Materials and Methods

3.7.1 Strains

The construction of the ancestral strains used in this chapter (listed in Table 3.2) was described elsewhere (Chapter 2). Briefly, we replaced the stator genes of *Escherichia coli* K12 MG1655 with stator genes from bacteria across the bacterial kingdom. The endogenous stator genes on the genome were replaced by a *rpsL-Neo* cassette [66], after which *rpsL* was replaced by *rpsL150*. Foreign stator genes were expressed from the low-copy plasmid pBAD33 [67]. This plasmid also contained a chloramphenicol-resistance gene. During the selection experiment (Chapter 2), the ancestral strains evolved (improved) motility. As the evolved strains are essentially similar to their respective ancestral strains (except for mutations), they are not listed in Table 3.2.

Table 3.2. Strains used in this chapter. NGS: Next Generation Sequencing.

Strain ID	Host strain	Stator genes on plasmid pBAD33	Donor strain stator genes	Number of pools for NGS
Strain 08	<i>Escherichia coli</i> K12 str. MG1655 Δ motAB (short Δ motAB)	<i>lafTU</i>	<i>Escherichia coli</i> O111:H-str. 11128	1
Strain 11	Δ motAB	<i>motPS</i>	<i>Bacillus pseudofirmus</i> OF4	1
Strain 13	Δ motAB	<i>lafTU</i>	<i>Photobacterium profundum</i> SS9	1
Strain 17	Δ motAB	<i>motAB</i>	<i>Listeria monocytogenes</i> EGD-e	2
Strain 18	Δ motAB	<i>motPS</i>	<i>Bacillus megaterium</i> DSM319	3
Strain 29	Δ motAB	<i>motAB</i> "1"	<i>Rhodospirillum centenum</i> SW	2

3.7.2 Preparation DNA samples for Next Generation Sequencing

In order to identify strain specific mutations after whole-genome re-sequencing by Next Generation Sequencing (NGS), we decided to pool endpoints per strain (Table 3.2). Endpoints were defined as genotypes that did not evolve improved motility during the selection experiment (Chapter 2). We were to sequence 10 pools and we distributed our endpoints accordingly. Endpoints were grown overnight separately (O/N; 5 μ l from -80°C stock; 16-17h; 250rpm; 37°C ; orbit diameter 2.5cm) in 10 ml liquid LB medium (20 g/l LB powder) supplemented with 0.0025% chloramphenicol (w/v). Next, the optical density at 600 nm (OD600) was measured in order to pool similar numbers of cells per endpoint. Here, we assumed that the relation between optical density and numbers of cells did not change during evolution. For instance, 1 ml of a culture with an OD600 of 3.5 was mixed with 2 ml of a culture with an OD600 of 1.75. After we pooled the endpoints, we used 1 ml of this pool to isolate genomic DNA (Promega Wizard Genomic DNA purification kit) and 4.5 ml of the same pool to isolate plasmid DNA (Machery-Nagel Nucleospin[®] Plasmid QuickPure). The rest of the sample of pooled endpoints was stored at -20°C , in case DNA isolation needed to be repeated. DNA concentration of both genomic DNA and plasmid DNA was determined using the Nanodrop (Thermo Scientific) and quality was determined by gel electrophoresis. Equimolar amounts of genomic DNA and plasmid DNA were mixed to yield final DNA samples containing $>3 \mu\text{g}$ DNA. These samples were analysed by ServiceXS B.V. using Illumina HiSeq (paired-end, 100 read length).

3.7.3 High Resolution Melt Analysis

In order to validate NGS data, High Resolution Melt Analysis (HRMA) was performed. HRMA is based on PCR-product melting in combination with a double-stranded DNA (dsDNA) binding fluorescent dye and is increasingly used as a fast and cheap method for genotyping in viruses and both pro- and eukaryotes [68–77]. As the melting temperature of dsDNA depends on the specific base composition of that strand, a mutation can easily be found by amplifying a specific piece of DNA by PCR, followed by comparing characteristics of the melt curve of the PCR product to the melt curve of a reference PCR product [71]. Here, we performed HRMA on the Illumina Eco Real-time PCR system and we used Solis BioDyne HOT FIREPol[®] EvaGreen[®] HRM Mix (No ROX). This HRM mix contained the dsDNA-binding fluorescent dye, DNA polymerase, ATP and nucleotides.

O/N cultures of a focal strain and a reference strain were diluted in sterile MilliQ to an OD600 of 0.170 and a sample from these dilutions, together with primers appropriate for the focal mutation,

was added to the HRM mix. The reference strain was always the ancestral strain (Table 3.2) of the focal strain and HRMA was performed in triplicate per strain. Mutations A→T, T→A, G→C and C→G in the focal strain could not be distinguished from the reference strain by comparing melt curves. In order to identify these mutations, the sample of the focal strain was made heterogeneous before performing HRMA. This was done by mixing cells from the focal strain with cells from the reference strain. After melting of the PCR product, the absolute fluorescence units of the melt curves in the *melt region* (from ±1°C before melting to ±1°C after melting) were normalized automatically to relative fluorescence units (RFU) by the Illumina ECO software. This was done to set the fluorescence before melting to 100% RFU and after melting to 0% RFU. In order to perform statistical analysis on the HRMA data (section 3.7.4), the Euclidian distance between the normalized melt curve of a standard (we used the average of the measurements of the reference strain) and all samples was calculated for every 0.1°C in the *melt region* [75]. The square root of the sum of these Euclidian distances was then used to test for significant differences between the reference strain and the focal strain. False positives could be identified in all cases, together with visual inspection of the melt curve (corroborated by Sanger sequencing). Together, mutations were identified by significant differences compared with the reference strain in Euclidian distance and by visual inspection of the melt curve.

3.7.4 Statistical analysis

Statistical analysis was only applied to identify mutations by HRMA. We used Student's *t*-tests and compensated for multiple comparisons by applying the Benjamini-Hochberg procedure [78].

3.8 Literature

- 1 Pallen M & Matzke N (2006) From The Origin of Species to the origin of bacterial flagella. *Nat. Rev. Microbiol.* **4**, 784–790.
- 2 Liu R & Ochman H (2007) Stepwise formation of the bacterial flagellar system. *Proc. Natl. Acad. Sci.* **104**, 7116–7121.
- 3 Mulkidjanian A & Makarova K (2007) Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. *Nat. Rev. Microbiol.* **5**, 892–899.
- 4 Dolezal P, Likic V, Tachezy J & Lithgow T (2006) Evolution of the molecular machines for protein import into mitochondria. *Science* **313**, 314–8.
- 5 Clements A, Bursac D, Gatsos X, Perry AJ, Civciristov S, Celik N, Likic VA, Poggio S, Jacobs-Wagner C, Strugnell RA & Lithgow T (2009) The reducible complexity of a mitochondrial molecular machine. *Proc. Natl. Acad. Sci.* **106**, 15791–15795.
- 6 Archibald JM, Logsdon Jr. JM & Doolittle WF (2000) Origin and Evolution of Eukaryotic Chaperonins: Phylogenetic Evidence for Ancient Duplications in CCT Genes. *Mol. Biol. Evol.* **17**, 1456–1466.
- 7 Gabaldón T, Rainey D & Huynen M (2005) Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I). *J. Mol. Biol.* **348**, 857–70.
- 8 Seidl MF & Schultz J (2009) Evolutionary flexibility of protein complexes. *BMC Evol. Biol.* **9**, 155.

- 9 Finnigan GC, Hanson-Smith V, Stevens TH & Thornton JW (2012) Evolution of increased complexity in a molecular machine. *Nature* **481**, 360–4.
- 10 Plotkin JB & Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42.
- 11 Goodman DB, Church GM & Kosuri S (2013) Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science* **342**, 475–479.
- 12 Blair DF (2003) Flagellar movement driven by proton translocation. *FEBS Lett.* **545**, 86–95.
- 13 Sowa Y & Berry RM (2008) Bacterial flagellar motor. *Q. Rev. Biophys.* **41**, 103–32.
- 14 Chevance FF V & Hughes KT (2008) Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.* **6**, 455–65.
- 15 Wadhams GH & Armitage JP (2004) Making sense of it all: bacterial chemotaxis. *Nat. Rev. Mol. Cell Biol.* **5**, 1024–37.
- 16 Adler J (1966) Chemotaxis in bacteria. *Science* **153**, 708–716.
- 17 Chun S & Parkinson J (1988) Bacterial motility: membrane topology of the Escherichia coli MotB protein. *Science* **239**, 276–278.
- 18 Khan S, Dapice M & Reese TS (1988) Effects of mot Gene Expression on the Structure of the Flagellar Motor at the Marine Biological Laboratory. *J. Mol. Biol.* **202**, 575–584.
- 19 de Mot R & Vanderleyden J (1994) The C-terminal sequence conservation between OmpA-related outer membrane proteins and MotB suggests a common function in both Gram-positive and Gram-negative bacteria, possibly in the interaction of these domains with peptidoglycan. *Mol. Microbiol.* **12**, 333–334.
- 20 Koebnik R (1995) Proposal for a peptidoglycan-associating alpha-helical motif in the C-terminal regions of some bacterial cell-surface proteins. *Mol. Microbiol.* **16**, 1269–1270.
- 21 Zhou J, Sharp LL, Tang HL, Lloyd SA, Billings S, Braun TF & Blair DF (1998) Function of Protonatable Residues in the Flagellar Motor of Escherichia coli: a Critical Role for Asp 32 of MotB. *J. Bacteriol.* **180**, 2729–2735.
- 22 Braun TF, Al-mawsawi LQ, Kojima S & Blair DF (2004) Arrangement of Core Membrane Segments in the MotA / MotB Proton-Channel Complex of Escherichia coli. *Biochemistry* **43**, 35–45.
- 23 Kojima S & Blair DF (2004) Solubilization and Purification of the MotA / MotB Complex of Escherichia coli. *Biochemistry* **43**, 26–34.
- 24 Zhou J & Blair DF (1997) Residues of the Cytoplasmic Domain of MotA Essential for Torque Generation in the Bacterial Flagellar Motor. *J. Mol. Biol.* **273**, 428–439.
- 25 Lloyd SA & Blair DF (1997) Charged Residues of the Rotor Protein FliG Essential for Torque Generation in the Flagellar Motor of Escherichia coli. *J. Mol. Biol.* **266**, 733–744.
- 26 Zhou J, Lloyd SA & Blair DF (1998) Electrostatic interactions between rotor and stator in the bacterial flagellar motor. *Proc. Natl. Acad. Sci.* **95**, 6436–6441.
- 27 Silverman M, Matsumura P & Simon M (1976) The identification of the mot gene product with Escherichia coli-lambda hybrids. *Proc. Natl. Acad. Sci.* **73**, 3126–3130.
- 28 Blair DF & Berg HC (1988) Restoration of Torque in Defective Flagellar Motors. *Science* **242**, 1678–1681.

- 29 Lind PA, Tobin C, Berg OG, Kurland CG & Andersson DI (2010) Compensatory gene amplification restores fitness after inter-species gene replacements. *Mol. Microbiol.* **75**, 1078–89.
- 30 Perica T, Chothia C & Teichmann S a (2012) Evolution of oligomeric state through geometric coupling of protein interfaces. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8127–32.
- 31 Hashimoto K & Panchenko AR (2010) Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20352–20357.
- 32 Davis BH, Poon AFY & Whitlock MC (2009) Compensatory mutations are repeatable and clustered within proteins. *Proc. Biol. Sci.* **276**, 1823–7.
- 33 Poon AFY & Chao L (2006) Functional Origins of Fitness Effect-Sizes of Compensatory mutations in the DNA Bacteriophage ϕ X174. *Evolution (N. Y.)*. **60**, 2032–2043.
- 34 Garza A, Harris-Haller L, Stoebner R & Manson M (1995) Motility protein interactions in the bacterial flagellar motor. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 1970–4.
- 35 Garza A, Biran R, Wohlschlegel J & Manson M (1996) Mutations in motB suppressible by changes in stator or rotor components of the bacterial flagellar motor. *J. Mol. Biol.* **258**, 270–85.
- 36 Hosking ER & Manson MD (2008) Clusters of charged residues at the C terminus of MotA and N terminus of MotB are important for function of the Escherichia coli flagellar motor. *J. Bacteriol.* **190**, 5517–21.
- 37 Passmore SE, Meas R & Marykwas DL (2008) Analysis of the FliM/FliG motor protein interaction by two-hybrid mutation suppression analysis. *Microbiology* **154**, 714–24.
- 38 Brockhurst MA, Colegrave N & Rozen DE (2011) Next-generation sequencing as a tool to study microbial evolution. *Mol. Ecol.* **20**, 972–80.
- 39 Illumina (2012) *Quality Scores for Next-Generation Sequencing*.
- 40 Agresti A & Coull BA (1998) Approximate Is Better than “Exact” for Interval Estimation of Binomial Proportions. *Am. Stat.* **52**, 119–126.
- 41 Jovelin R, Dunham JP, Sung FS & Phillips PC (2009) High nucleotide divergence in developmental regulatory genes contrasts with the structural elements of olfactory pathways in caenorhabditis. *Genetics* **181**, 1387–97.
- 42 Barrier M, Robichaux RH & Purugganan MD (2001) Accelerated regulatory gene evolution in an adaptive radiation. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10208–13.
- 43 Duboule D & Wilkins AS (1998) The evolution of “bricolage”. *Trends Genet.* **14**, 54–9.
- 44 Blank D, Wolf L, Ackermann M & Silander OK (2014) The predictability of molecular evolution during functional innovation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3044–9.
- 45 Rogozin IB, Makarova KS, Natale DA, Spiridonov AN, Tatusov RL, Wolf YI, Yin J & Koonin E V (2002) Congruent evolution of different classes of non-coding DNA in prokaryotic genomes. *Nucleic Acids Res.* **30**, 4264–4271.
- 46 Roujeinikova A (2008) Crystal structure of the cell wall anchor domain of MotB , a stator component of the bacterial flagellar motor : Implications for peptidoglycan recognition. , 1–6.
- 47 Reboul CF, Andrews DA, Nahar MF, Buckle AM & Roujeinikova A (2011) Crystallographic and molecular dynamics analysis of loop motions unmasking the peptidoglycan-binding site in stator protein MotB of flagellar motor. *PLoS One* **6**, e18981.

- 48 O'Brien E & Bennet P (1972) Structure of Straight Flagella from a Mutant Salmonella. *J. Mol. Biol.* **70**, 133–152.
- 49 Berg HC (2004) *E. coli in motion* Springer.
- 50 Morimoto Y V, Nakamura S, Kami-ike N, Namba K & Minamino T (2010) Charged residues in the cytoplasmic loop of MotA are required for stator assembly into the bacterial flagellar motor. *Mol. Microbiol.* **78**, 1117–29.
- 51 Morimoto Y V, Nakamura S, Hiraoka KD, Namba K & Minamino T (2013) Distinct roles of highly conserved charged residues at the MotA-FliG interface in bacterial flagellar motor rotation. *J. Bacteriol.* **195**, 474–81.
- 52 Krogh A & Rapacki K (2013) <http://www.cbs.dtu.dk/services/TMHMM/>. .
- 53 Hosking ER, Vogt C, Bakker EP & Manson MD (2006) The Escherichia coli MotAB proton channel unplugged. *J. Mol. Biol.* **364**, 921–37.
- 54 Blair DF, Kim DY & Berg HC (1991) Mutant MotB Proteins in Escherichia coli. *J. Bacteriol.* **173**.
- 55 Marykwas DL & Berg HC (1996) A mutational analysis of the interaction between FliG and FliM, two components of the flagellar motor of Escherichia coli. *J. Bacteriol.* **178**, 1289–94.
- 56 Lenski RE, Rose MR, Simpson SC, Tadler SC, Naturalist TA & Dec N (1991) Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. *Am. Nat.* **138**, 1315–1341.
- 57 Rainey PB & Travisano M (1998) Adaptive radiation in a heterogeneous environment. *Nature* **32**, 69–72.
- 58 Hall AR, Griffiths VF, MacLean RC & Colegrave N (2010) Mutational neighbourhood and mutation supply rate constrain adaptation in Pseudomonas aeruginosa. *Proc. Biol. Sci.* **277**, 643–50.
- 59 Burch C & Chao L (2000) Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* **406**, 625–628.
- 60 Chen S, Beeby M, Murphy GE, Leadbetter JR, Hendrixson DR, Briegel A, Li Z, Shi J, Tocheva EI, Müller A, Dobro MJ & Jensen GJ (2011) Structural diversity of bacterial flagellar motors. *EMBO J.* **30**, 2972–81.
- 61 Gerrish PJ & Lenski RE (1998) The fate of competing beneficial mutations in an asexual population. *Genetica* **102-103**, 127–44.
- 62 Moore FB, Rozen DE & Lenski RE (2000) Pervasive compensatory adaptation in Escherichia coli. *Proc. Biol. Sci.* **267**, 515–22.
- 63 Open WetWare Escherichia coli Codon Usage. .
- 64 Vimberg V, Tats A, Remm M & Tenson T (2007) Translation initiation region sequence preferences in Escherichia coli. *BMC Mol. Biol.* **8**, 100.
- 65 Takekawa N, Kojima S & Homma M (2014) Contribution of many charged residues at the stator-rotor interface of the Na⁺-driven flagellar motor to torque generation in Vibrio alginolyticus. *J. Bacteriol.* **196**, 1377–85.
- 66 Heermann R, Zeppenfeld T & Jung K (2008) Simple generation of site-directed point mutations in the Escherichia coli chromosome using Red(R)/ET(R) Recombination. *Microb. Cell Fact.* **7**, 14.
- 67 Guzman LM, Belin D, Carson MJ, Beckwith J, Guzman L, Belin D & Carson MJ (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130.
- 68 Fitzcharles EM (2012) Rapid discrimination between four Antarctic fish species, genus Macrourus, using HRM analysis. *Fish. Res.* **127-128**, 166–170.

- 69 Nery J (2010) Amplicon Genotyping and Methylation Analysis by HRM. .
- 70 Robertson T, Bibby S, O'Rourke D, Belfiore T, Agnew-Crumpton R & Noormohammadi AH (2010) Identification of Chlamydial species in crocodiles and chickens by PCR-HRM curve analysis. *Vet. Microbiol.* **145**, 373–9.
- 71 Vossen RHAM, Aten E, Roos A & den Dunnen JT (2009) High-resolution melting analysis (HRMA): more than just sequence variant screening. *Hum. Mutat.* **30**, 860–6.
- 72 Lee ASG, Ong DCT, Wong JCL, Siu GKH & Yam W-C (2012) High-resolution melting analysis for the rapid detection of fluoroquinolone and streptomycin resistance in *Mycobacterium tuberculosis*. *PLoS One* **7**, e31934.
- 73 Kalthoff D, Beer M & Hoffmann B (2013) High resolution melting analysis: rapid and precise characterisation of recombinant influenza A genomes. *Virolog. J.* **10**, 284.
- 74 Tanaka M, Takahashi J, Hirayama F & Tani Y (2011) High-resolution melting analysis for genotyping Duffy, Kidd and Diego blood group antigens. *Leg. Med. (Tokyo)*. **13**, 1–6.
- 75 Hjelmsø MH, Hansen LH, Bælum J, Feld L, Holben WE & Jacobsen CS (2014) High Resolution Melt analysis for rapid comparison of bacterial community composition. *Appl. Environ. Microbiol.* **80**, 3568–3575.
- 76 Chen D, Wang Y-Y, Chuai Z-R, Huang J-F, Wang Y-X, Liu K, Zhang L-Q, Yang Z, Shi D-C, Liu Q, Huang Q & Fu W-L (2014) High-resolution melting analysis for accurate detection of BRAF mutations: a systematic review and meta-analysis. *Sci. Rep.* **4**, 4168.
- 77 Tindall EA, Petersen DC, Woodbridge P, Schipany K & Å VMH (2009) Assessing High-Resolution Melt Curve Analysis for Accurate Detection of Gene Variants in Complex DNA. .
- 78 Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300.

3.9 Supplementary information

Table 3.3 to 3.8 are shown on the next pages.

Table 3.3. Mutations in regulatory and structural genes found in strain 11. Second part of the table focusses on BFM genes. (Potential) physiological function indicates the potential involvement of genes in affecting the proton motive force (PMF) or peptidoglycan layer only.

Gene	Structural / regulatory	Biological process	Number of mutations	(Potential) physiological function
<i>baeS</i>	Regulatory	Regulation / signal transduction	1	
<i>zraS</i>	Regulatory	Regulation / signal transduction	8	
<i>sspA</i>	Regulatory	Transcription regulation	1 (1bp del)	
<i>ltaE</i>	Structural	Amino acid synthesis	1	
<i>ltaK</i>	Structural	DNA recombination	1	
<i>narY</i>	Structural	Electron transport	1	Affecting PMF
<i>aegA</i>	Structural	Electron transport	1	Affecting PMF
<i>fabB</i>	Structural	Membrane synthesis	1	
<i>yciM</i>	Structural	Membrane synthesis	1	
<i>acrE</i>	Structural	Transport	1	Affecting PMF (ecocyc)
<hr/>				
			Mutation	Function
<i>motS</i>	Structural	Motility	Q52P E149K	ion- and peptidoglycan-binding part of stator

Table 3.4. Mutations in regulatory and structural genes found in strain 17. Second part of the table focusses on BFM genes and chemotaxis genes. (*Potential*) *physiological function* indicates the potential involvement of genes in affecting the proton motive force (PMF) or peptidoglycan layer only.

Pool	Gene	Structure/regulatory	Biological process	Number of mutations	(Potential) physiological function
1,2	<i>cbpM</i>	Regulatory	Regulation	1	
1	<i>atpC</i>	Structural	ATP synthesis	1	Affecting PMF
1	<i>atpA</i>	Structural	ATP synthesis	2	Affecting PMF [Terashima]
1	<i>ycgB</i>	Structural	Cell division	1	
1	<i>coaE</i>	Structural	CoA metabolism	1	
1	<i>accA</i>	Structural	Membrane synthesis	1	
1	<i>fabH</i>	Structural	Membrane synthesis	1. [1bp del]	
1	<i>yjbB</i>	Structural	Membrane synthesis	1. [18bp del]	
1	<i>accD</i>	Structural	Membrane synthesis	1	
1	<i>dacA</i>	Structural	Peptidoglycan synthesis	1	Affecting peptidoglycan
2	<i>ppsA</i>	Structural	Pyruvate metabolism	1	
1	<i>tujA</i>	Structural	Protein synthesis	1	
1	<i>lepB</i>	Structural	Protein processing	2	
1,2	<i>mHA</i>	Structural	Transport	1. [1bp del]	Affecting PMF

Pool	Gene	Structure/regulatory	Biological process	Mutation	Function
2	<i>motA</i>	Structural	Motility	I166M	Rotor-interacting part of stator
1				T26A	
1				T26I	
1,2				S39I	
2	<i>motB</i>	Structural	Motility	S39R	Ion- and peptidoglycan-binding part of stator
1				A43E	
1				A52T	
2				F53V	
2				F53L	
1				E99A	
1				G166C	
2	<i>flgG</i>	Structural	Motility	G194S	Stator-interacting part of rotor
1				L275F	
2				Q280R	
1	<i>flc</i>	Structural	Motility	K20N	Protein making up the flagellum
1	<i>cheB</i>	Structural	Motility	S12L	Modifies receptor of chemotaxis cascade
1	<i>cheW</i>	Structural	Motility	S12R	Linker between receptor and CheA

Table 3.5. Mutations in regulatory and structural genes found in strain 18. Second part of the table (next page) focusses on BFM genes and chemotaxis genes. (*Potential*) *physiological function* indicates the potential involvement of genes in affecting the proton motive force (PMF) or peptidoglycan layer only.

Pool	Gene	Structural/regulatory	Biological process	Number of mutations	(Potential) physiological function
1	<i>quuQ</i>	Regulatory	Transcription regulation	1 (1bp ins)	
1	<i>slxA</i>	Regulatory	Transcription regulation	1	
1	<i>sspA</i>	Regulatory	Transcription regulation	1	
1	<i>yfeD</i>	Regulatory	Transcription regulation	1	
2,3	<i>ydhB</i>	Regulatory	Transcription regulation	2	
2	<i>phoP</i>	Regulatory	Transcription regulation	1	
2	<i>purR</i>	Regulatory	Transcription regulation	1	
1	<i>baxS</i>	Regulatory	Regulation / signal transduction	1	
1	<i>yfjK</i>	Regulatory	?	1	
1,3	<i>yeeG</i>	Structural	Antibiotic resistance/cell division	2 (of which 1 4bp del)	
3	<i>atpD</i>	Structural	ATP synthesis	1	Interacts with FIG [Zharbiv]
1	<i>mhpA</i>	Structural	Carbohydrate metabolism	1	
1	<i>amyA</i>	Structural	Carbohydrate metabolism	1	
1	<i>nema</i>	Structural	Electron transport	1	Affecting PMF
1,2	<i>lpp</i>	Structural	Membrane integrity	2	
1,2,3	<i>senA</i>	Structural	Membrane integrity	4	
2	<i>wcaF</i>	Structural	Membrane synthesis	1	
1,2	<i>dacA</i>	Structural	Peptidoglycan synthesis	2 (1bp del)	Affecting peptidoglycan
1	<i>ltdB</i>	Structural	Peptidoglycan synthesis	1 (1bp del)	Affecting peptidoglycan
2	<i>ybS</i>	Structural	Peptidoglycan synthesis	1	Affecting peptidoglycan
1	<i>infB</i>	Structural	Protein synthesis	1	
2	<i>tufA</i>	Structural	Protein synthesis	1	
1	<i>rlnM</i>	Structural	rRNA processing	1	
2	<i>yjE</i>	Structural	Transport	1	Affecting PMF (ecocyc)

Table 3.5 Continued.

Pool	Gene	Structural/regulatory	Biological process	Mutation	Function
1				E91K	
1				G123R	
1,2	<i>motA</i>	Structural	Motility	E125K	Rotor-interacting part of stator
1,2				L234R	
1,2,3				Q237K	
2				S239Y	
2				D25E	
1,3				F29L	
2	<i>motB</i>	Structural	Motility	L36R	Ion- and peptidoglycan-binding part of stator
2				L54Q	
2				A169V	
1				A171T	
2				Q252K	
2	<i>fljG</i>	Structural	Motility	E253K	Stator-interacting part of rotor
1				Q280R	
2				E49K	
2				V58M	
1,2	<i>fljM</i>	Structural	Motility	A161V	CheY-interacting part of rotor
2				N195D	
1				L10F	
2	<i>fljC</i>	Structural	Motility	T157I	Flagellar filament structural protein
2				K166R	
1	<i>cheB</i>	Structural	Motility	V311L	Modifies receptor of chemotaxis cascade

Table 3.6. Mutations in regulatory and structural genes found in strain 08. Second part of the table focusses on BFM genes. *(Potential) physiological function* indicates the potential involvement of genes in affecting the proton motive force (PMF) or peptidoglycan layer only.

Gene	Structural/regulatory	Biological process	Number of mutations	(Potential) physiological function
<i>ycsG</i>	Structural	Antibiotic resistance/cell division (septation protein)	1	
<i>atpA</i>	Structural	ATP-synthesis	1	Affecting PMF [Terashima]
			Mutation	Function
			E87K	
			Q90K	
<i>lafT</i>	Structural	Motility	A134S	Rotor-interacting part of stator
			L284R	
<i>flG</i>	Structural	Motility	Q280R	Stator-interacting part of rotor

Table 3.7. Mutations in regulatory and structural genes found in strain 13. Second part of the table focusses on BFM genes and chemotaxis genes. (*Potential*) *physiological/function* indicates the potential involvement of genes in affecting the proton motive force (PMF) or peptidoglycan layer only.

Gene	Structural/regulatory	Biological process	Number of mutations	(Potential) physiological function
<i>proC</i>	Regulatory	Transcription regulation	1	
<i>phiA</i>	Regulatory	Transcription regulation	1	Affecting PMF (through dehydrogenase regulation)
<i>rpoD</i>	Regulatory	Transcription regulation	1	
<i>yocC</i>		?	1	
<i>yqjB</i>		?	1	
<i>psrD</i>		?	1	
<i>yjE</i>	Structural	Stress response	1	
<i>rraB</i>	Structural	DNA replication	1	
<i>rrdF</i>	Structural	DNA replication	1	
<i>sarA</i>	Structural	Membrane integrity	1	
<i>mrcB</i>	Structural	Peptidoglycan synthesis	2 (of which 1.1bp cel)	Affecting peptidoglycan
<i>prc</i>	Structural	Protein catabolism	3	
<i>ybhI</i>	Structural	Transport	1	Affecting PMF
<i>ydhK</i>	Structural	Transport	1	Affecting PMF
<i>ydfN</i>	Structural	Transport	1	Affecting PMF
<i>ygdN</i>	Structural	Transport	1	Affecting PMF (scoyC)

Gene	Structural/regulatory	Biological process	Mutation	Function
			I174L	
			Y199H	
<i>lgfJ</i>	Structural	Motility	S236A	Ion- and peptidoglycan-binding part of stator
			V252A	
			Y271D	
			D150G	
<i>flg</i>	Structural	Motility	E177V	Stator-interacting part of rotor
			R190H	
			S279P	
			G68E	
<i>flhM</i>	Structural	Motility	A161V	CheY-interacting part of rotor
			N173D	
			N173I	
			N201H	
<i>cheA</i>	Structural	Motility	D541E	Modifies the effector CheY of chemotaxis cascade
			A35V	
<i>cheB</i>	Structural	Motility	M59L	Modifies receptor of chemotaxis cascade
			G298R	
<i>Isr</i>	Structural	Motility	E391G	Receptor of chemotaxis cascade

Table 3.8. Mutations in regulatory and structural genes found in strain 29. Second part of the table (next page) focusses on BFM genes and chemotaxis genes. (*Potential*) *physiological function* indicates the potential involvement of genes in affecting the proton motive force (PMF) or peptidoglycan layer only.

Pool	Gene	Structural/regulatory	Biological process	Number of mutations	(Potential) physiological function
1	<i>abrB</i>	Regulatory	Transcription regulation	1	
1	<i>rpoD</i>	Regulatory	Transcription regulation	2	
2	<i>rno</i>	Regulatory	Transcription regulation	1	
2	<i>envZ</i>	Regulatory	Regulation / signal transduction	2	
2	<i>ompR</i>	Regulatory	Regulation / signal transduction	3 (of which 1.16bp del)	
2	<i>zraR</i>	Regulatory	Regulation / signal transduction	1	
1	<i>roxA</i>		?	1	
1	<i>araA</i>	Structural	Carbohydrate metabolism	1	
1	<i>alsK</i>	Structural	Carbohydrate metabolism	1	
2	<i>fsh</i>	Structural	Protein catabolism	1	Degradation stator proteins
2	<i>cadB</i>	Structural	Transport	1	Affecting PMF (ecocyc)
1	<i>pitA</i>	Structural	Transport	1	Affecting PMF (ecocyc)
1	<i>proP</i>	Structural	Transport / Proton symport	1	Affecting PMF
2	<i>ubiX</i>	Structural	Ubiquinone biosynthesis	1	Affecting PMF through ubiquinone
1	<i>thiL</i>	Structural	Vitamin B1 synthesis	1	

Table 3.8 Continued.

Pool	Gene	Structural/regulatory	Biological process	Mutation	Function
1,2				N2N	
1,2	<i>motA</i>	Structural	Motility	L3M I5I	Rotor-interacting part of stator
1				G6G	
	<i>motB</i>	Structural	Motility	Y64N	Ion- and peptidoglycan-binding part of stator
1				I133L	
2	<i>flgG</i>	Structural	Motility	E224Q E49K	Stator-interacting part of rotor
1				E59V	
2				T100A	
1	<i>fljM</i>	Structural	Motility	P174T	CheY-interacting part of rotor
2				N201H	
2				L80L	
1	<i>fljH</i>	Structural	Motility		Part of flagellar export apparatus
1	<i>flgF</i>	Structural	Motility	V237G	Component of cell-proximal portion of rod
2	<i>cheB</i>	Structural	Motility	A41T	Modifies receptor of chemotaxis cascade
2	<i>cheW</i>	Structural	Motility	V98A	Linker between receptor and CheA
Pool	Locus	(Potential) role			
1,2	6bp to <i>MotA</i>	Shine dalgarno of <i>MotA</i>			
1,2	P15A Origin	Origin of replication plasmid			

4 Mutational and phenotypic trajectories of foreign component integration in the bacterial flagellar motor

Régis C.E. Flohr, Stefanie Luik, Marcel Langelaan, Erwin van Rijn, Francesco Pedaci, Hubertus J.E. Beaumont. *In preparation for publication.*

4.1 Abstract

Experimental evidence of the underlying molecular mechanisms of compositional evolution is scarce. Compositional evolution, which is evolution by the step-wise addition of pre-existing components to simpler assemblies, is suggested as a mechanism by which protein complexes originated, innovated and adapted. In the previous chapters, we showed that incompatible foreign components could be evolutionarily integrated by mutations along multiple selectively-accessible adaptive trajectories. In this chapter, we examine the detailed phenotypic and genetic step-wise changes that occurred in a number of focal lineages for a sub-set of the mutational targets. The results reveal the location and temporal order of mutations in flagellar motor genes and in genes encoding chemotactic signalling. Different stators were integrated along different trajectories as did replicate events of integration of the same stator. This genetic divergence was paralleled by divergent step-wise changes in a range of motility related phenotypes. Together, these and additional observations reveal a detailed snapshot of the pathways that were available for compositional evolution in the BFM.

4.2 Introduction

Compositional evolution, evolution by the addition of pre-existing protein components to an already-existing protein complex [1], is suggested to have played a major role in the evolution of protein complexes [2–9]. However, experimental evidence on the underlying mechanisms is scarce. The previous two chapters of this thesis described the results of an experiment that examined how incompatible homologous components were integrated into the *Escherichia coli* (*E. coli*) cell. Results showed that foreign components could be integrated by mutations (evolutionary integration) along different adaptive trajectories. This conclusion was based on several lines of evidence, namely phenotypic differences between evolutionary lineages (number and length of evolutionary lineages per strain), phenotypic differences between genotypes selected during the evolution of increased motility (population swimming speed) and mutational differences between pools of replicate selection lines. However, direct insight into the step-wise phenotypic and genetic changes that occurred within these replicate lineages was still lacking.

In this chapter, we intended to sample the diversity in both mutational and phenotypic changes that occurred after each step in the adaptive trajectories and that facilitated stator integration. In Chapter 2, we replaced stator genes of the bacterial flagellar motor (BFM) of *E. coli* with 22 different stator genes (foreign stators) from across the bacterial kingdom, which resulted in 22 strains harbouring a chimeric BFM (cBFM strains). The BFM is the main structure for bacterial motility, it consists of more than 20 different protein components [10] and can be divided into four major structural features. One of these structures, the stator complex, consists of two proteins. These proteins are MotA and MotB [11,12] and have distinct functions. MotB is involved in anchoring of the BFM to the cell wall [13,14] and in ion binding [15], whereas MotA couples the translocation of ions from outside to inside the cell to flagellar rotation and torque generation [10,16–18]. As such, the stator complex is essential for BFM-mediated motility. In eight of the cBFM strains we constructed, the foreign stators were compatible immediately. This was determined by investigating chemotaxis-driven motility of populations of cells in semi-solid agar. In three of these instances, we observed the possibility for the evolution of improved motility: genotypes with improved motility escaped from their ancestral population and formed wedge-shaped populations of cells (flares). In three other strains, we first observed a possibility for the evolution of compatibility (strains evolved motility), followed by the evolution of improved motility. The overall characteristics of the underlying mutations were discussed in Chapter 3 and a key finding was that individual stators could be integrated along different trajectories and that stators from different donor strains were integrated in part by identical mutations. Furthermore, the mutations we discovered were clustered with most mutations being located in the stator genes.

The main objective of this chapter was to investigate the diversity in adaptive trajectories, both phenotypic and mutational, underlying the evolutionary integration of foreign components. We achieved this by investigating in more detail a subset of the evolutionary lineages that were generated during the main selection experiment (Chapter 2). Each of these lineages consisted of genotypes isolated from consecutive flares that formed on the soft agar plates. Our characterization of these genotypes included the examination of population phenotypes (competitive motility against a fluorescently-labelled, wildtype strain), single-cell phenotypes (swimming speed, percentage non-motile cells in the population, cell size), single-BFM phenotypes (number of tumbling cells as a

measure of tumbling frequency) and the order of mutations in *focal genes*. The focal genes included genes of the BFM and genes of the chemotaxis signalling pathway (hereafter, *BFM genes* and *chemotaxis genes*, respectively).

The goal of these more detailed genetic and phenotypic analyses was to address, among others, the following questions. Can we observe multiple mutational changes per lineage as indicated by the occurrence of consecutive flaring events and bulk genetic analysis? If multiple changes occurred in the lineages, what are their temporal dynamics and order? Is the same component sometimes mutated multiple times? Do individual trajectories contain mutations in multiple BFM components? What is the pattern of change in lower-level motility related phenotypes that is associated with the step-wise increases in motility? Answering these questions provides direct insight into the diversity of the step-wise adaptive changes that are accessible for incorporation of a foreign protein. In addition, it will provide insight into the underlying (molecular) mechanisms that constrain this evolutionary process.

Our results revealed multi-step trajectories that differ between replicate selection lines and cBFM strains. In some cases trajectories contained multiple mutations in a single BFM component and sequential changes in multiple BFM components. The pattern of evolutionary change in lower-level motility related phenotypes varied between replicate lineages and cBFM strains. Together, these and additional observations revealed detailed snapshots of the pathways that were available for compositional evolutionary tinkering in the BFM.

4.3 Results

When evolution of improved motility occurred during the selection experiment, i.e. when a flare was generated, a single genotype was isolated from this flare (hereafter called *strain*; *Materials and Methods*). This strain was subsequently used to 1) start the next round of selection (therefore consecutive strains are directly related) and 2) perform phenotypic and mutational characterizations on. Preliminary screening of 50 strains isolated from a single flare, showed that all strains were more motile than their immediate ancestor, although phenotypic differences were observed (supplementary information of Chapter 2). These phenotypic differences were not investigated in more detail, but might be a result of differences in initial population size.

4.3.1 Phenotypic adaptive trajectories at the level of the population and single cell

Our selection experiment favoured cells capable of producing flares. As such, we selected for cells that were capable of escaping from their ancestral population in semi-solid agar by evolving an increased ability to colonize unoccupied regions of the agar plate. In this section, we quantified the competitive ability of strains from consecutive flares (consecutive strains) to colonize the swimming plate (*competitive motility*) and a number of lower-level, single-cell phenotypes that are closely associated with this ability. The single-cell phenotypes we examined were single-cell swimming speed, cell size and the percentage of non-motile cells in the population.

Competitive motility was measured against a fluorescently-labelled variant of the reference strain harbouring wildtype *E. coli* stators and is defined as the ratio of velocities of the population

fronts of the competing strains (*Materials and Methods*). Because flare formation represents an increase in the ability to spread out in semi-solid agar, genotypes isolated from flares were expected to have an increased competitive motility. It is, however, important to point out that this fitness assay does not recapitulate the exact environment in which the genotypes escaped from the population front of their ancestors to form a flare.

The single-cell phenotypes examined in this chapter were used by us previously (Chapter 2) in order to investigate phenotypic effects of foreign stators that were compatible immediately. The motility of populations of cells in nutrient-rich, semi-solid agar is caused by chemotaxis, which is the ability of cells to sense and respond to chemical gradients [19–21]. When this gradient is a nutrient-gradient, a population of cells will follow the gradient to higher concentrations, while at the same time consuming nutrients and thus steepening the gradient. The generation and steepness (we assume a sigmoidal shape [22]) of this nutrient gradient is influenced by the consumption rate of the nutrient by a population of cells, which in turn is dependent on several factors, including single-cell consumption rate, single-cell growth rate or population size. As the percentage of non-motile cells in a population influences the size of the population that performs chemotaxis, it might affect the generation and steepness of the attractant gradient and thus affect population motility. Single-cell swimming speed, on the other hand, determines the ability of following the moving attractant-gradient by the population and might influence population motility also. As our cBFM strains had a severely impaired or completely lacking population motility, this phenotype was expected to have been modified during the selection experiment. The pore size of semi-solid agar is in the order of the cell width of *E. coli* [23–26]. Therefore, cell size might influence the ability of cells to move through the pores and affect population motility. Lastly, cellular tumbling, which is caused by the temporary reversal of the BFM rotation, allows a cell to re-orient itself (e.g. [27,28]) and is known to effect motility of cells in agar [29].

The selection experiment yielded many evolutionary lineages. To investigate differences both within and between cBFM strains, we selected a short and long evolutionary lineage from strains 11 and 18, both of which harboured a sodium stator that was compatible immediately.

4.3.1.1 Short evolutionary lineage of cBFM strain 11

The short evolutionary lineage of cBFM strain 11 consisted of only three strains, including the ancestral strain. First, we investigated the competitive motility of strains against a fluorescently-labelled strain harbouring wildtype stators (strain 01). Although the competitive motility increased for consecutive strains, strain 11.1c.2a was still significantly slower than strain 01 (Figure 4.1A; Benjamini-Hochberg-corrected Student's *t*-tests; $n = 3$; $p < 0.001$ for all comparisons). This result indicated that the genotypes isolated from consecutive flares had an improved motility compared with their direct ancestors. This is consistent with the idea that the selection experiment selected for cells with a higher fitness.

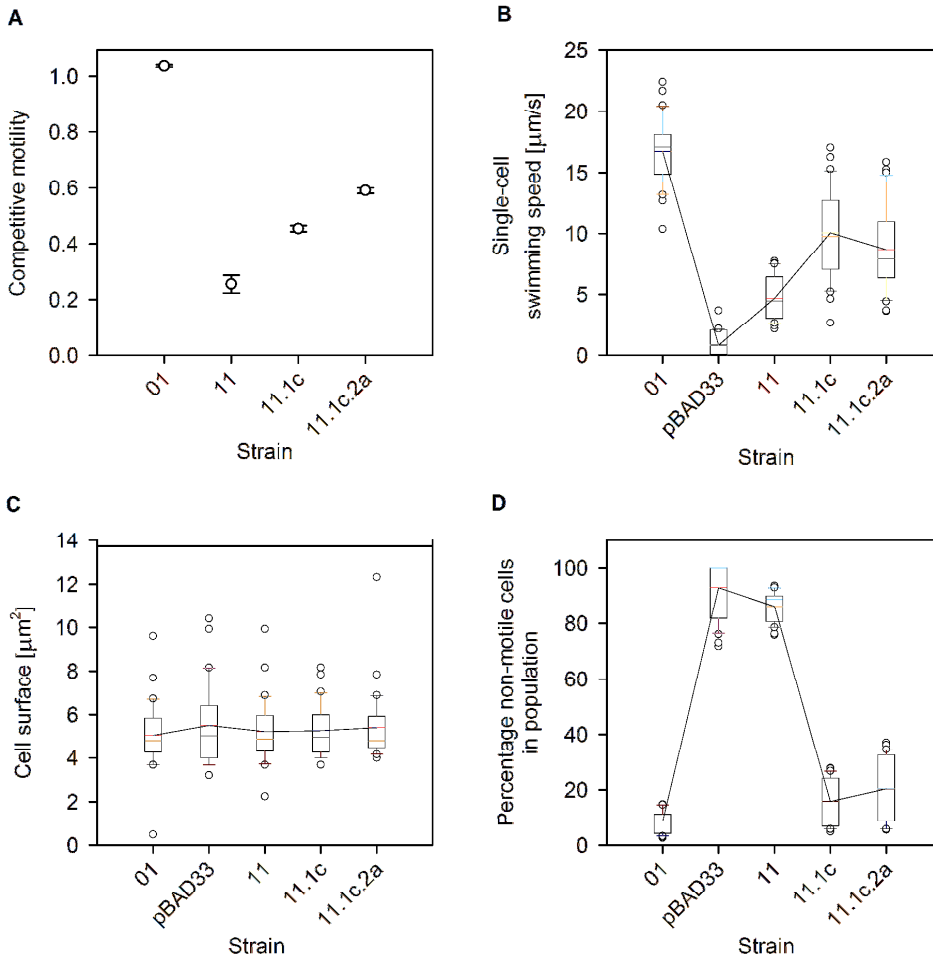


Figure 4.1. Population and single-cell phenotypes in a short evolutionary lineage of cBFM strain 11. **A.** Competitive motility against a fluorescently-labelled version of the reference strain harbouring wildtype *E. coli* stator genes. Competitive motility is defined as the ratio of expansion velocities between competing strains (*Materials and Methods*). Bars indicate standard deviation from mean. $n = 3$. **B.** Single-cell swimming speed measured in a liquid medium. **C.** Surface of single cells. **D.** The percentage of non-motile cells in the population. **B-D.** $n = 30$. Boxes indicate the middle 50% of data points, with whiskers above and below indicating the 90th and 10th percentile, respectively. Circles indicate data points outside the 10 and 90th percentile. The horizontal lines in boxes that are interconnected by a line indicate means. The other horizontal line in a box indicates the median. pBAD33 is a statorless strain. See text for statistics.

In order to investigate the underlying mechanisms of the increase in competitive motility, we examined single-cell phenotypes of the consecutive strains. While the competitive motility increased for consecutive strains, the single-cell swimming speed only increased significantly for cells of strain 11.1c, but not for cells of strain 11.1c.2a (Figure 4.1B; Benjamini-Hochberg-corrected Mann-Whitney *U*-tests, $p = 0.001$ and 0.376 , respectively). Similar results, with respect to significant differences, were found for the percentage of non-motile cells in populations (Figure 4.1D; Benjamini-Hochberg-corrected Mann-Whitney *U*-tests, $p = 0.001$ for 11.1c and 0.412 for 11.1c.2a). For both phenotypes,

we found strain 11.1c.2a to be significantly different from strain 01 (Benjamini-Hochberg-corrected Mann-Whitney U -tests, $p < 0.01$ for both comparisons). As a change in cell size might have been involved in the generation of flares, we also analysed this phenotype (Figure 4.1C). However, cell size did not change for any of the strains.

The single-cell swimming speed of cBFM strain 11 was significantly higher than that of the statorless and therefore non-motile strain pBAD33. The single-cell swimming speed measured for the latter strain was due to drift and/or Brownian motion (observation) and these processes might have influenced the speed of other strains also. However, assuming the influence is equal for all strains, differences between strains do not change. Together with the high percentage of non-motile cells in the population, a low single-cell swimming speed may explain the atypical motility (ATM) observed in semi-solid agar for cBFM strain 11. We defined ATM as a type of motility in semi-solid agar characterised by dots radiating from the point of inoculation (after 24h), which represented small bacterial colonies growing in the agar surrounding the inoculation site. Previously [17,30,31], it was proposed that this type of motility was the results of a very small fraction of motile cells in a population that usually produced daughter cells that were non-motile. However, a low tumbling frequency and low torque (which results in low single-cell swimming speed) might also underlie ATM (Chapter 2).

4.3.1.2 Long evolutionary lineage of cBFM strain 11

The second and longer evolutionary lineage of cBFM strain 11 that we investigated led to strain 11.1j.2a.3a.4a. Figure 4.2A shows an increasing competitive motility against the fluorescently-labelled strain harbouring the wildtype stators, although not all strains were more motile than their direct ancestor (*Discussion*). For instance, strain 11.1j was not more motile than strain 11 and strain 11.1j.2a.3a was not more motile than strain 11.1j.2a (Benjamini-Hochberg-corrected Student's t -tests; $n = 3$; $p = 0.06$ and 0.55 respectively).

The single-cell swimming speeds showed different dynamics than found for competitive motility (Figure 4.2B). Strain 11.1j was significantly faster than strain 11 (Benjamini-Hochberg-corrected Mann-Whitney U -test; $n = 30$, $p < 0.001$), after which none of the consecutive strains was faster than their direct ancestor (Benjamini-Hochberg-corrected Mann-Whitney U -tests; $n = 30$, $p > 0.05$ for all comparisons). This was similar to what we observed for the short evolutionary lineage of cBFM strain 11.

The percentage of non-motile cells in a population (Figure 4.2D) showed dynamics that were different from either of the phenotypes discussed so far. We observed a steady decrease in the percentage of non-motile cells for consecutive strains, but only some differences were significant. These were the differences between strains 11 and 11.1j and between strains 11.1j.2a and 11.1j.2a.3a (Benjamini-Hochberg-corrected Mann-Whitney U -tests; $n = 30$, $p < 0.01$ for both comparisons).

Strain 11.1j.2a.3a.4a was significantly different from strain 01 for all three discussed phenotypes (Benjamini-Hochberg-corrected Mann-Whitney U -tests; $p < 0.03$ for all comparisons). Cell surface was not significantly different for any of the strains.

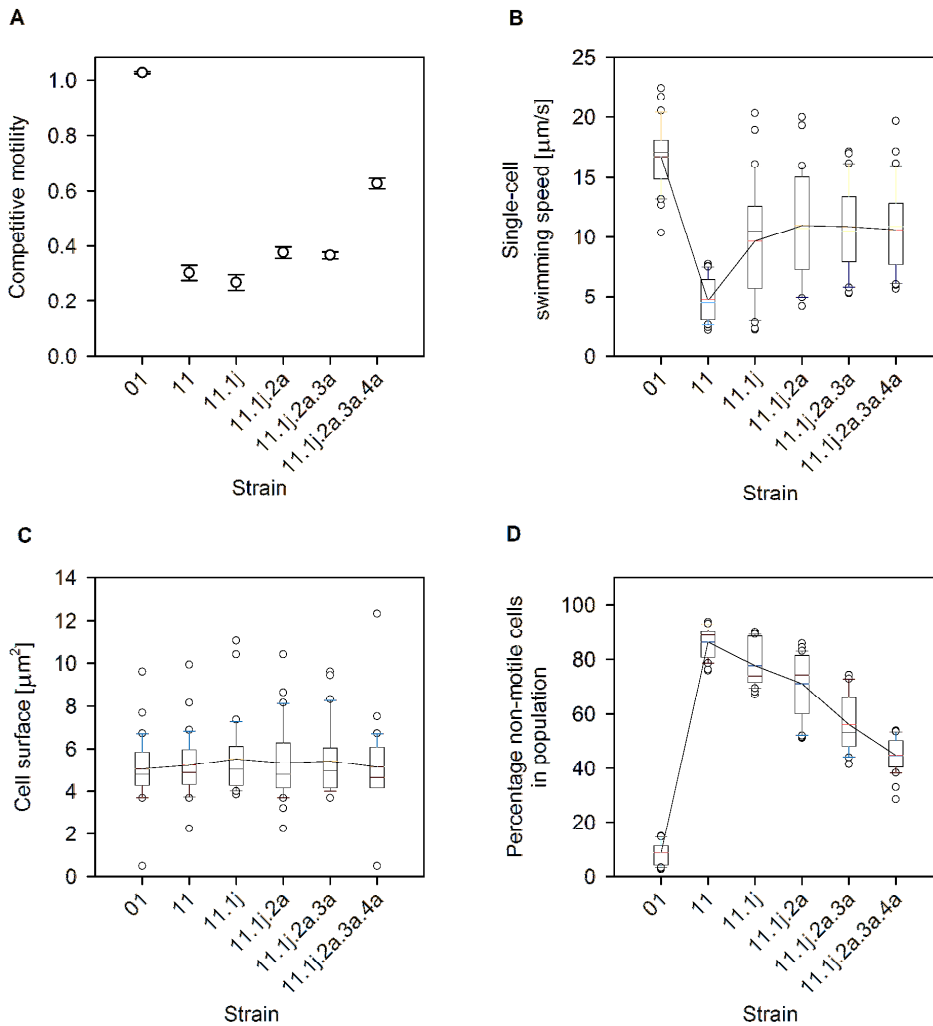


Figure 4.2. Population and single-cell phenotypes in a long evolutionary lineage of cBFM strain 11. **A.** Competitive motility against a fluorescently-labelled version of the reference strain harbouring wildtype *E. coli* stator genes. Competitive motility is defined as the ratio of expansion velocities between competing strains (*Materials and Methods*). Bars indicate standard deviation from mean. $n = 3$. **B.** Single-cell swimming speed measured in a liquid medium. **C.** Surface of single cells. **D.** The percentage of non-motile cells in the population. **B-D.** $n = 30$. Boxes indicate the middle 50% of data points, with whiskers above and below indicating the 90th and 10th percentile, respectively. Circles indicate data points outside the 10 and 90th percentile. The horizontal lines in boxes that are interconnected by a line indicate means. The other horizontal line in a box indicates the median. See text for statistics.

Although the competitive motility of the endpoints (the final strain in an evolutionary lineage) of the short and long evolutionary lineage of strain 11 was similar, the phenotypic adaptive trajectories that led to them were different. Not only did the trajectories differ in length, but also were the dynamics of phenotypic change different. For instance, the short lineage showed a steady increase in competitive motility, whereas the long lineage showed smaller steps initially. The most noticeable

difference however, was the decrease in percentage of non-motile cells in a population. The short lineage showed a sudden decrease, whereas the long lineage showed a linear decrease by small steps.

4.3.1.3 Short evolutionary lineage of cBFM strain 18

No significant increase in competitive motility in the short evolutionary lineage of cBFM strain 18 was observed until strain 18.1h.2a.3a, which was significantly more motile than its direct ancestor (Figure 4.3A; Benjamini-Hochberg-corrected Mann-Whitney U -test; $n = 30$, $p < 0.001$). On the other hand, strain 18.1h.2a was significantly less motile than strain 18.1h (Benjamini-Hochberg-corrected Mann-Whitney U -test; $n = 30$, $p = 0.023$; see *Discussion*). Similar results were found for single-cell swimming speeds: strain 18.1h.2a.3a showed a significantly higher single-cell swimming speed than its direct ancestor (Figure 4.3B; Benjamini-Hochberg-corrected Mann-Whitney U -test; $n = 30$, $p = 0.001$).

The percentage of non-motile cells (Figure 4.3D) was significantly different between all consecutive strains (Benjamini-Hochberg-corrected Mann-Whitney U -tests; $n = 30$, $p < 0.05$ for all comparisons). However, a counter intuitive result was obtained for strain 18.1h, which showed a significantly higher percentage of non-motile cells in the population than strain 18. An explanation for this result might be that other factors caused this genotype to evolve by natural selection and that the low percentage of non-motile cells was the result of associated pleiotropic effects.

Cell surface was not significantly different for any of the strains. Similar to cBFM strain 11, cBFM strain 18 performed ATM on semi-solid agar. A higher single-cell swimming speed and a lower percentage of non-motile cells than the statorless strain pBAD33 (Benjamini-Hochberg-corrected Mann-Whitney U -tests; $n = 30$, $p < 0.005$ for both comparisons) might explain the motility observed.

Strain 18.1h.2a.3a was significantly different from strain 01 for all phenotypes except cell surface (Benjamini-Hochberg-corrected Mann-Whitney U -tests; $p < 0.04$ for all comparisons).

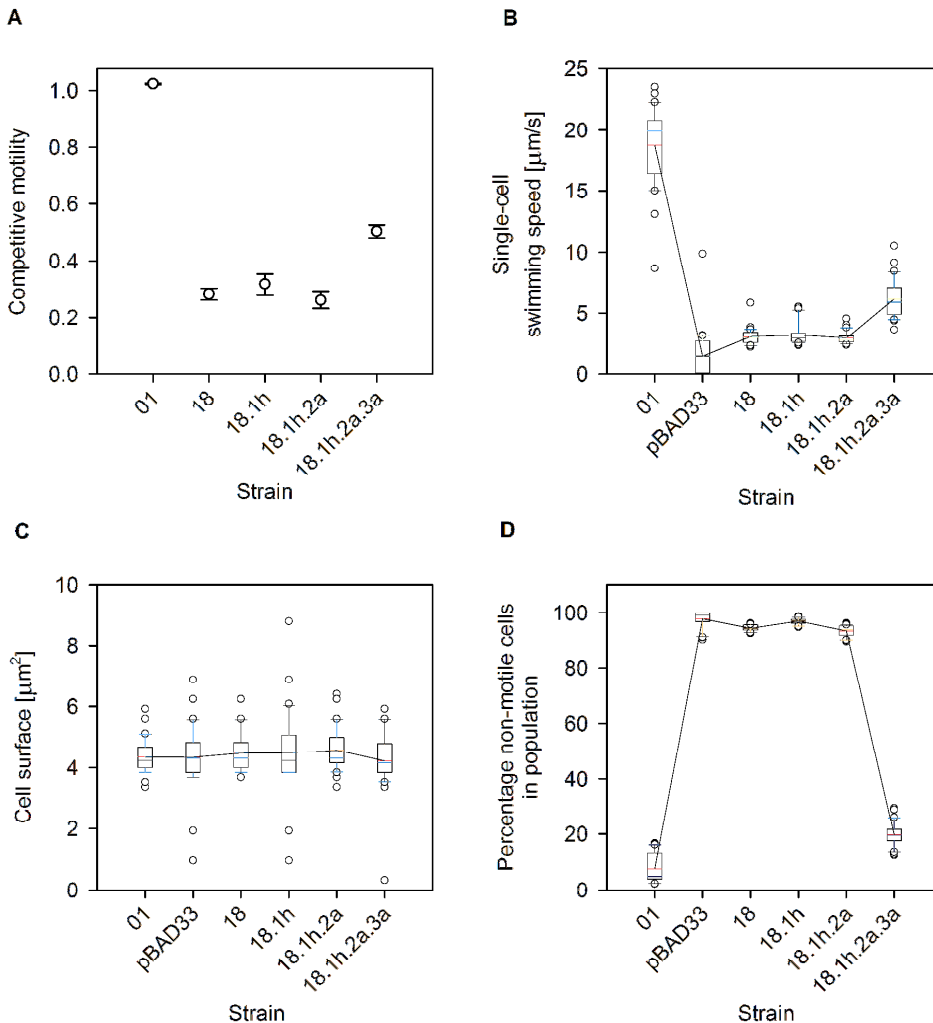


Figure 4.3. Population and single-cell phenotypes in a short evolutionary lineage of cBFM strain 18. **A.** Competitive motility against a fluorescently-labelled version of the reference strain harbouring wildtype *E. coli* stator genes. Competitive motility is defined as the ratio of expansion velocities between competing strains (*Materials and Methods*). Bars indicate standard deviation from mean. $n = 3$. **B.** Single-cell swimming speed measured in a liquid medium. **C.** Surface of single cells. **D.** The percentage of non-motile cells in the population. **B-D.** $n = 30$. Boxes indicate the middle 50% of data points, with whiskers above and below indicating the 90th and 10th percentile, respectively. Circles indicate data points outside the 10 and 90th percentile. The horizontal lines in boxes that are interconnected by a line indicate means. The other horizontal line in a box indicates the median. pBAD33 is a statorless strain. See text for statistics.

4.3.1.4 Long evolutionary lineage of cBFM strain 18

With six consecutive flares, the long evolutionary lineage of cBFM strain 18 was twice as long as the short evolutionary lineage. The competitive motility between consecutive strains did not increase significantly until strain 18.1d.2a.3a.4a.5a, which was significantly more motile than its ancestor and significantly less motile than the strain it generated (Figure 4.4A; Benjamini-Hochberg-corrected

Student's *t*-test; $n = 30$, $p < 0.005$ for both comparisons). Furthermore, strain 18.1d.2a.3a.4a.5a.6a was significantly less motile than strain 01 (Benjamini-Hochberg-corrected Mann-Whitney *U*-test; $n = 3$, $p < 0.001$ for both comparisons).

Similar trends were observed for both the single-cell swimming speed (Figure 4.4B) and the percentage of non-motile cells in a population (Figure 4.4C): strain 18.1d.2a.3a.4a.5a was significantly different from its direct ancestor (Benjamini-Hochberg-corrected Mann-Whitney *U*-test; $n = 30$, $p < 0.01$ for both comparisons). However, strain 18.1d.2a.3a.4a.5a.6a was not significantly different from strain 18.1d.2a.3a.4a.5a in both cases, although differences were observed (Benjamini-Hochberg-corrected Mann-Whitney *U*-test; $n = 30$, $p = 0.085$ and 0.137 respectively). Furthermore, strain 18.1d.2a.3a.4a.5a.6a did also not differ significantly from strain 01 for both single-cell swimming speed and the percentage of non-motile cells in a population. We attribute this lack of significance despite obvious differences, to the fact that non-parametric statistical tests were performed. When performing these tests, the actual data were converted to ranks and these ranks were compared between groups. As such, large differences in actual data became smaller in ranked data. Cell surface was not significantly different for any of the strains.

The short and long evolutionary lineages of strain 18 showed similar phenotypic dynamics, both at the level of the population as well as at the level of single cells. The clearest difference, besides length, was found in the percentage of non-motile cells in the population. The short lineage showed a one-step drop from 92.5% to 19.7%, whereas the long lineage showed a two-step drop from 96.4% to 66% to 34.4%.

Similar to strain 11, these results indicated that different phenotypic adaptive-trajectories to improved motility were available for strains harbouring identical, foreign stators. However, in contrast to strain 11, in which genotypes isolated from primary flares already showed phenotypic changes (especially at the single-cell level), in strain 18 we found phenotypic changes to occur near the end of the evolutionary lineages. These results indicated that different stators resulted in different accessible adaptive-trajectories.

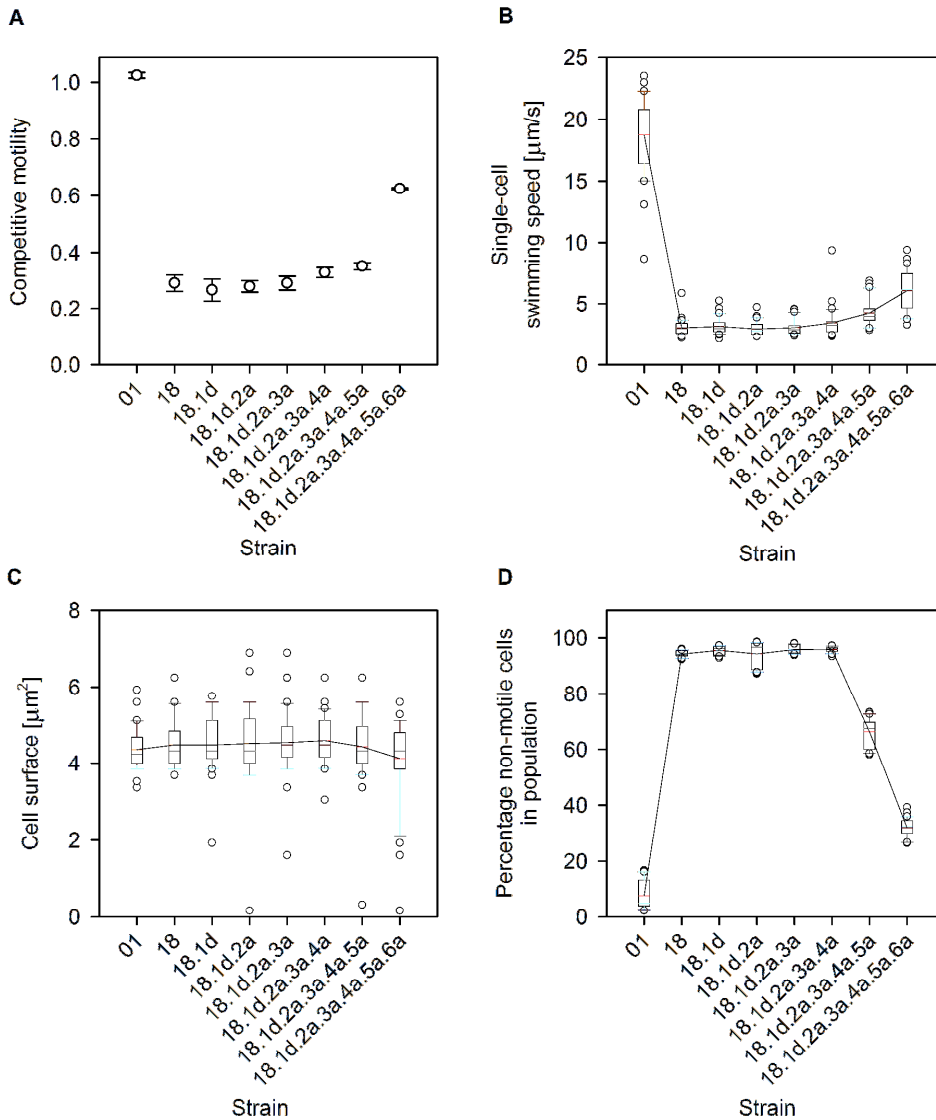


Figure 4.4. Population and single-cell phenotypes in a long evolutionary lineage of cBFM strain 18. **A.** Competitive motility against a fluorescently-labelled version of the reference strain harbouring wildtype *E. coli* stator genes. Competitive motility is defined as the ratio of expansion velocities between competing strains (*Materials and Methods*). Bars indicate standard deviation from mean. $n = 3$. **B.** Single-cell swimming speed measured in a liquid medium. **C.** Surface of single cells. **D.** The percentage of non-motile cells in the population. **B-D.** $n = 30$. Boxes indicate the middle 50% of data points, with whiskers above and below indicating the 90th and 10th percentile, respectively. Circles indicate data points outside the 10 and 90th percentile. The horizontal lines in boxes that are interconnected by a line indicate means. The other horizontal line in a box indicates the median. See text for statistics.

4.3.2 Mutational adaptive trajectories: resolving the order of mutations

In the previous sections, we investigated the dynamics of population and single-cell phenotypes for several evolutionary lineages and the results showed differences in adaptive trajectories both within cBFM strains and between cBFM strains. Because the ultimate source of phenotypic changes can be found as mutations in the DNA, we were also interested in mutational adaptive trajectories. In order to find these mutations, we performed whole-genome re-sequencing on the endpoints that had evolved during the selection experiment (Chapter 3). This provided us with an overview of the mutations in the final genotypes of an evolutionary lineage and to resolve the sequence of mutations that had accumulated in a single evolutionary lineage, we used High Resolution Melt Analysis (HRMA) [32]. These results would not only demonstrate potential diversity in accessible mutational adaptive-trajectories, but would also shed light on the molecular mechanisms of evolution of (improved) motility.

Because of the large number of mutational targets (Chapter 3), we restricted our analysis to genes of the BFM and the chemotaxis signalling pathway (i.e. focal genes). These focal genes were expected to have a high probability of being involved in compatibilization and increasing motility, in case of incompatible stators and compatible stators, respectively. Furthermore, we focussed on two strains that were selected based on their evolutionary dynamics. While strain 08 showed short evolutionary lineages on average, strain 18 required more steps on average.

4.3.2.1 cBFM strain 08

cBFM strain 08 yielded nine endpoints originating from five primary flares (Figure 4.5). In these nine endpoints we found seven mutations in genes in total (Chapter 3), five of which were located in the focal genes. Four of these mutations were found in stator gene *lafT* and the fifth mutation was located in rotor gene *fliG*, of which the protein product is the interaction partner of LafT [10]. The result of ordering these five mutations by HRMA in all evolved strains is shown in Figure 4.5.

All strains isolated from primary flares at least harboured the A134S mutation in the foreign stator LafT, which suggested that this might be the first mutation to fix in populations of cBFM strain 08. Three strains harboured an additional mutation. In two cases (8.1b and 8.1c), this involved the LafT L284R mutation. In the third case (8.1d) it involved the LafT E87K mutation (*Discussion*).

The evolutionary lineages that originated from strain 8.1e did not show additional mutations until the strains isolated from the quaternary flares. Although the strains that evolved from strain 8.1e.2a.3a all harboured the LafT L284R mutation, the situation differed compared with strains 8.1b and 8.1c. Because stator genes were expressed from a low copy plasmid [33], multiple copies of the stator genes (alleles) were present in the cell. Theoretically it would thus be possible to have only non-mutated alleles (homozygous wildtype), only mutated alleles (homozygous mutant) or a combination of both wildtype and mutant alleles (heterozygous). While in strains 8.1b and 8.1c the L284R mutation was homozygous, the mutation was heterozygous in offspring of strain 8.1e.2a.3a. Furthermore, the nature of heterozygosity was different for different strains. The ratio between wildtype and mutant alleles was in favour of the wildtype allele in strain 8.1e.2a.3a.4a and the strain that evolved from it, but the ratio was in favour of the mutated allele in strain 8.1e.2a.3a.4b.

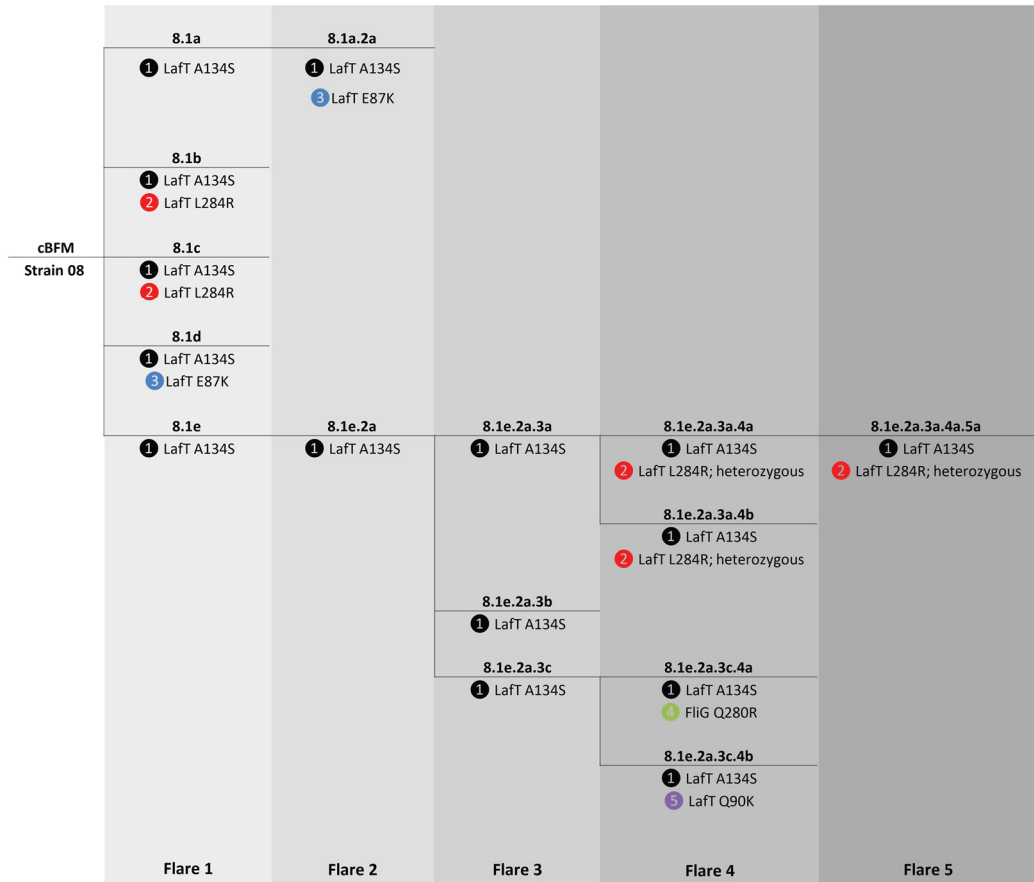


Figure 4.5. The sequence of mutations in focal genes in evolutionary lineages of cBFM strain 08. The heterozygous mutations found in 8.1e.2a.3a.4a and 8.1e.2a.3a.4a.5a were in favour of the wildtype allele. In 8.1e.2a.3a.4b, heterozygosity was in favour of the mutant allele. See text for details. Different colours indicate different mutations.

Two unique mutations in addition to LaFT A134S were found in strains 8.1e.2a.3c.4a and 8.1e.2a.3c.4b. In the latter strain it involved an additional mutation in LaFT, namely Q90K, while in the former strain it involved mutation FlIG Q280R.

For two evolutionary lineages, we also measured the competitive motility of the strains (Figure 4.6). Although we cannot claim causality of mutations (for this we need to reconstitute individual mutations), expressed in terms of the contribution of mutations to increases in competitive motility in semi-solid agar, it does indicate a possible effect.

Strain 8.1a harboured the LaFT A134S mutation, which caused ATM in semi-solid agar (see Chapter 5). However, competitive motility did not increase significantly compared with strain 08, which was non-motile (Benjamini-Hochberg-corrected Student's *t*-test; $n=3$, $p=0.053$). Strain 8.1a.2a showed a significantly higher competitive motility than its direct ancestor (Benjamini-Hochberg-corrected *t*-test; $n=3$, $p<0.001$). The mutation (LaFT E87K) we found in this strain causes a charge

reversal near the Laft-FliG interaction site, thereby potentially being a functional mutation with large effects.

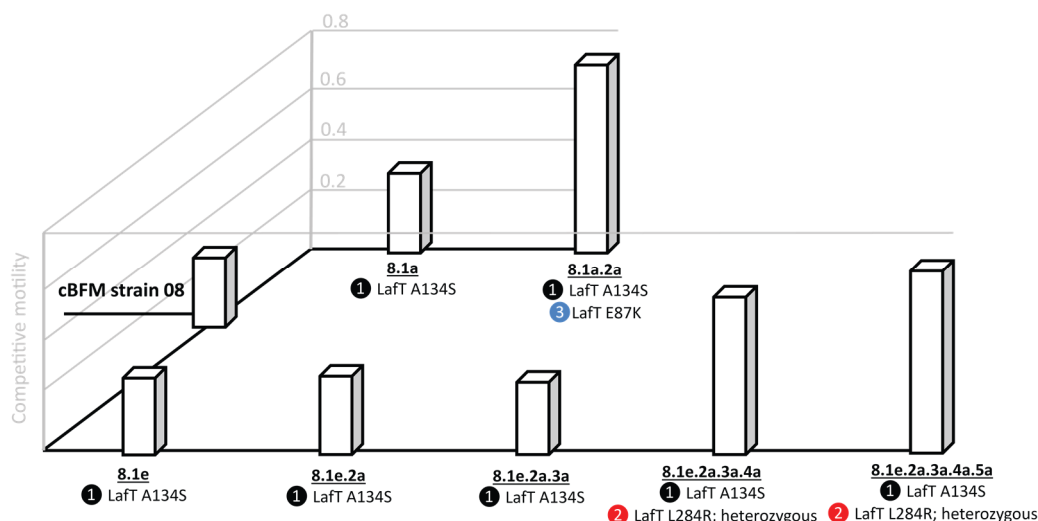


Figure 4.6. Competitive motility accompanying mutations in Laft in two evolutionary lineages of cBFM strain 08. Competitive motility was measured against a fluorescently-labelled version of the reference strain harbouring wildtype *E. coli* stator genes. Competitive motility is defined as the ratio of expansion velocities between competing strains (*Materials and Methods*). $n = 3$.

Similar to strain 8.1a, no significant increase in competitive motility was discovered for strain 8.1e. Although additional mutations in other genes might be present in strains 8.1e.2a and 8.1e.2a.3a (NGS identified additional mutations), no changes in competitive motility were found (Benjamini-Hochberg-corrected t -tests; $n=3$, $p>0.05$ for all comparisons). Accompanying the heterozygous Laft L284R mutation in strain 8.1e.2a.3a.4a, we found the competitive motility to increase significantly compared with its direct ancestor (Benjamini-Hochberg-corrected t -test; $n=3$, $p<0.001$). Although strain 8.1e.2a.3a.4a.5a was significantly more motile than its direct ancestor (Benjamini-Hochberg-corrected t -test; $n=3$, $p<0.001$), no additional mutations were discovered. Moreover, the heterozygosity of the Laft L284R mutation was unchanged (data not shown). This result implied the presence of additional mutations in this strain outside the focal genes.

In summary, all mutations that we discovered in the genotypes isolated from primary flares were located in the foreign stator Laft. Only in strain 8.1e.2a.3c.4a did a mutation occur in another BFM component, although mutations in genes not examined here had been identified by NGS (Chapter 3). Furthermore, these results showed directly that some evolutionary lineages took different mutational trajectories.

4.3.2.2 cBFM strain 18

Evolution of improved motility in cBFM strain 18 led to 14 primary flares, which yielded 38 endpoints. These endpoints were distributed over three pools, which were independently analysed

with NGS (Chapter 3). We selected pool 1 to investigate in more detail by HRMA, because of the many different focal genes that were mutated and because it contained less mutations than pool 2. Pool 1 contained 12 endpoints that evolved from 6 primary flares (1b, 1c, 1e, 1h, 1j and 1l). We discovered 27 unique mutations, of which 11 were located in 6 focal genes. The result of ordering these mutations by HRMA is shown in Figure 4.7.

cBFM strain 18	18.1b	18.1b.2a 3 FlIG Q280R	18.1b.2a.3a 3 FlIG Q280R 9 MotA E125K	
		18.1b.2b 4 MotA G123R heterozygous		
	18.1c	18.1c.2a 1 MotA Q237K	18.1c.2a.3a 1 MotA Q237K	
	18.1e	18.1e.2a 1 MotA Q237K	18.1e.2a.3a 1 MotA Q237K	
			18.1e.2a.3b 1 MotA Q237K 10 FlIG A171T	
		18.1e.2b 1 MotA Q237K	18.1e.2b.3a 1 MotA Q237K	
	18.1h	18.1h.2a 5 CheB V311L	18.1h.2a.3a 5 CheB V311L 11 MotA E91K	
			18.1h.2a.3b 5 CheB V311L 9 MotA E125K	
	18.1j	18.1j.2a 1 MotA Q237K	18.1j.2a.3a 1 MotA Q237K	
		18.1j.2b 1 MotA Q237K	18.1j.2b.3a 1 MotA Q237K	18.1j.2b.3a.4a 1 MotA Q237K
	18.1l	18.1l.2a 2 FlIM A161V 6 MotB F29L heterozygous		
		18.1l.2b 2 FlIM A161V 7 MotA L234R heterozygous 8 FlIC L10F		
	Flare 1	Flare 2	Flare 3	Flare 4

Figure 4.7. The sequence of mutations in focal genes in evolutionary lineages of cBFM strain 18. The heterozygous mutations found in 18.1b.2b, 18.1l.2a and 18.1l.2b were all in favour of the wildtype allele. Different colours indicate different mutations.

In contrast to the strains isolated from primary flares of cBFM strain 08, not all strains isolated from primary flares of cBFM strain 18 harboured mutations in the focal genes. Strains lacking these mutations were 18.1b, 18.1c and 18.1h. The results indicated that cBFM strain 18 was less constrained in the first mutational step compared with strain 08. Of the three strains that did harbour mutations in the focal genes, two (18.1e and 18.1j) harboured mutations in the foreign stator gene *motA* (similar to strain 08). The third strain harboured a mutation in *fliM*, a gene of which the protein product is part of the rotor and involved in switching of the direction of the BFM [34].

Strains with the MotA Q237K mutation mostly lacked additional mutations in focal genes. Only strain 18.1e.2a.3b gained an additional mutation in FlIG (A171T). In contrast, lineages without the MotA Q237K mutation showed diverse additional mutations. Some of these mutations were located in stator genes (both *motA* and *motB*), whereas other mutations were discovered in different BFM components (FlIG and FlIC) or the chemotaxis signalling pathway (CheB [27]).

Competitive motility of strains in the evolutionary lineage that led to endpoint 18.1h.2a.3a, was already discussed in section 4.3.1.3 (data reproduced as

Figure 4.8) and allowed us to investigate possible effects of mutations. Absence of a mutation in the focal genes in strain 18.1h was accompanied by the absence of an increase in competitive motility. The mutation CheB V311L was discovered in strain 18.1h.2a. As CheB is responsible for resetting the receptors to their unstimulated state in the chemotaxis signalling pathway [35], the mutation might be involved in chemotaxis sensitivity. Although this might positively affect motility in semi-solid agar, a significant lower competitive motility was found (*Discussion*). Strain 18.1h.2a.3a harboured mutation MotA E91K and this mutation was accompanied by an increase in competitive motility and single-cell swimming speed (Figure 4.3B) and a decrease in the percentage of non-motile cells in a population (Figure 4.3D). As this mutation causes a charge reversal near the MotA-FlIG interaction site, it might change the interaction between rotor and stator.

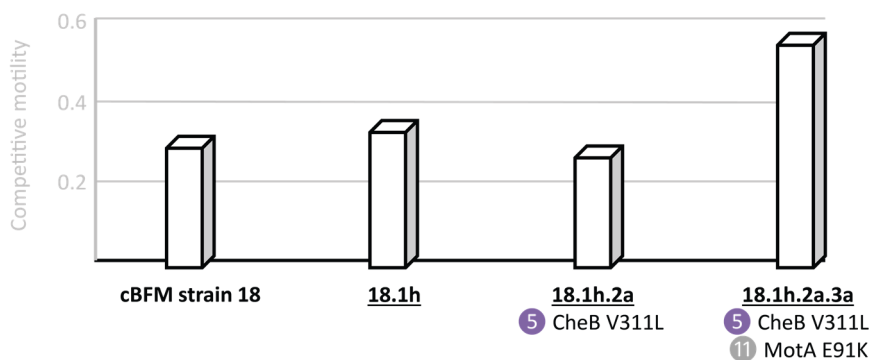


Figure 4.8. Competitive motility of strains in the evolutionary lineage leading to endpoint 18.1h.2a.3a. Competitive motility was measured against a fluorescently-labelled version of the reference strain harbouring wildtype *E. coli* stator genes. Competitive motility is defined as the ratio of expansion velocities between competing strains (*Materials and Methods*). $n = 3$.

In contrast to the results we found for cBFM strain 08, the mutations we discovered in strains isolated from primary flares in cBFM strain 18 were not constrained to the foreign stator genes. Furthermore, additional mutations were found in both BFM and chemotaxis signalling pathway

genes, indicating that the mutational possibilities to improve motility were also larger for cBFM strain 18 than for cBFM strain 08.

4.3.3 Adaptive trajectory at the level of a single BFM: a preliminary analysis in cBFM strain 17

For the evolutionary lineages we discussed so far, we investigated adaptive trajectories at the population level, single-cell level or nucleotide level. In this section, we present results from a phenotype measured at the level of individual BFM complexes. We focussed this analysis on cBFM strain 17, a strain that performed ATM initially. In Chapter 2, we suggested that ATM might be a result of low tumbling frequency of the cells. Consequently, evolution of improved motility in a strain that performs ATM might involve changes in the BFM that alter its switching frequency and thus fine-tune the tumbling frequency of the swimming cells.

Strain 17.1j.2a.3a had a significantly higher competitive motility against a fluorescently-labelled strain containing wildtype stators than strain 17 (Figure 4.9A; Benjamini-Hochberg-corrected Student's *t*-test; $n = 3$, $p < 0.001$), which indicated that evolution of improved motility had occurred in this evolutionary lineage. In order to examine whether changes in tumbling could underlie this observation, we investigated the frequency of BFM direction switching. The results revealed that strain 17.1j.2a.3a had BFMs that had a significantly higher switching frequency than its ancestor, strain 17 (Figure 4.9B; Benjamini-Hochberg-corrected χ^2 -test, $n = 122$, $p < 0.001$). This indicated that a change in tumbling frequency (assuming the fraction of cells that perform tumbling is correlated with tumbling frequency) might underlie evolution of improved motility in this evolutionary lineage. Next we investigated tumbling in cBFM strain 17 and how this compared to strain 01. We wondered whether cBFM strain 17 showed a lower competitive motility than strain 01 (Figure 4.9A; Benjamini-Hochberg-corrected Student's *t*-test; $n = 3$, $p < 0.001$) due to a lower tumbling frequency and whether the tumbling frequency of strain 17.1j.2a.3a was similar to WILDTYPE levels. Surprisingly, we found that the fraction of cells from strain 17 that performed tumbling was not significantly less than the fraction of cells from strain 01 (Figure 4.9B; Benjamini-Hochberg-corrected χ^2 -test, $n = 122$, $p = 0.142$), which indicated that other factors underlie the observed difference in competitive motility. Lastly, we found that strain 17.1j.2a.3a showed a significantly larger fraction of cells that performed tumbling than strain 01 (Figure 4.9B; Benjamini-Hochberg-corrected χ^2 -test, $n = 122$, $p < 0.001$).

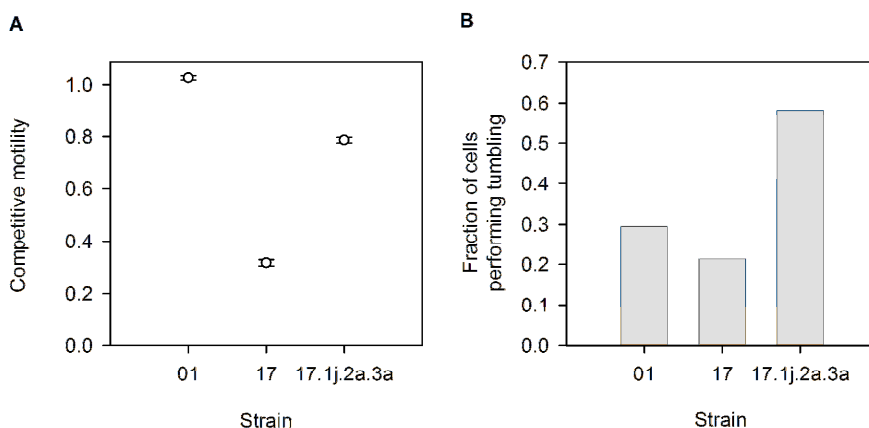


Figure 4.9. Competitive motility and the fraction of cells that showed at least one switching event for an evolutionary lineage of cBFM strain 17. **A.** Competitive motility against a fluorescently-labelled version of the reference strain harbouring wildtype *E. coli* stator genes. Competitive motility is defined as the ratio of expansion velocities between competing strains (*Materials and Methods*). Bars indicate standard deviation from mean. $n = 3$. **B.** Fraction of cells with BFM that switched direction at least once during a 15 second recording. $n = 122$. See text for statistics.

4.4 Discussion

In the previous chapters, we showed that different adaptive trajectories to improved motility were taken during the evolutionary integration of foreign stators. We achieved this by investigating the population swimming speeds of endpoints and the length and branching pattern of evolutionary lineages (Chapter 2) and by examining the overall pattern of mutations per cBFM strain (Chapter 3). In this chapter we investigated individual evolutionary lineages in more detail. We intended to directly examine differences in adaptive trajectories both within and between strains and examine the underlying genetic and phenotypic changes that facilitated functional integration of foreign stators into the BFM and host cell.

4.4.1 Diversity in phenotypic and mutational adaptive trajectories

Our results directly showed different adaptive trajectories both within cBFM strains and between cBFM strains in the evolution of improved motility. Not only did most of the trajectories we investigated differ in terms of the genetic changes, the dynamics of phenotypic change also differed, involving both the competitive fitness as well as the various lower-level phenotypes that were characterized. These results indicated that the selectively-accessible adaptive trajectories, both mutational and phenotypic, were diverse within a strain, but also depended on the identity of the foreign stator.

Two possible cases in which epistasis might have shaped the trajectories were observed. The first of these examples involved the LaFT L284R mutation in strain 08 (Figure 4.5). This mutation was heterozygous when it occurred in genotypes isolated from quaternary flares, but homozygous when it occurred in genotypes isolated from primary flares. Additional mutations might have resulted in the retention of heterozygosity (see also section 4.4.3). The second example of epistasis was found

in strain 18 (Figure 4.7). Here, occurrence of the MotA Q237K mutation appears to have limited the occurrence of additional mutations in BFM and chemotaxis genes. Future analysis of the longer-term evolutionary trajectory can reveal whether these different genetic trajectories impose differential constraints on the maximal fitness that can be reached or if they merely represent different routes to the same maximally attainable functionality.

Although the lower-level population phenotypes that we investigated might help to explain the evolution of improved motility in semi-solid agar, our results also indicated the presence of additional factors underlying evolution. For instance, in strain 11 we observed significant changes in competitive motility that were not accompanied by changes in the investigated single-cell phenotypes (Figure 4.1; strain 11.1c.2a). Because we discovered changes in tumbling in one evolutionary lineage of strain 17 (Figure 4.9), which is known to affect motility in semi-solid agar [29], this implies it might also underlie evolution in strain 11. Similarly, additional mutations besides the mutations in the focal genes were expected in some genotypes. Not only did we identify additional mutations in other genes (Chapter 3), but also did apparently similar genotypes show different population swimming speeds, e.g. 8.1a.2a and 8.1d (Figure 4.5 and Chapter 2). This latter result again indicated the presence of epistasis.

All five genotypes isolated from primary flares of cBFM strain 08 harboured mutations in the stator genes (Figure 4.5). In contrast, only two out of six genotypes isolated from primary flares of cBFM strain 18 showed a similar result (Figure 4.7). This difference might be explained by the difference in initial compatibility of the foreign stators. cBFM strain 18 harboured compatible stators, whereas cBFM strain 08 harboured incompatible stators. Our results suggested that mutations leading to compatibility were more constrained than mutations improving motility when the stators were compatible already, which we suggested before in a similar way. In Chapter 2 we suggested that with improved integration of the foreign stators, more mutations (at more loci) became available to improve motility further. We speculated that dominant conflicts that needed to be resolved first were the main cause for this effect.

We did not detect an increase in competitive motility in semi-solid agar between all consecutive strains and in one instance we even detected a decrease (Figure 4.3A). Three explanations, that are not mutually exclusive, might be given for these results. First, the method we used to measure competitive motility is not reliable at detecting differences at very low population swimming speeds. For instance, it is not possible to distinguish between ATM-performing and non-motile strains (Figure 4.6; strains 8.1a and 8.1e compared with strain 08). We suggest that this is true in general for small differences in motility. Second, non-transitive interactions occurred [36]. Non-transitive interactions may result in a lower or equal competitive motility of a focal strain against the reference strain harbouring wildtype *E. coli* stators, while this focal strain is actually more motile than its direct ancestor in a head-to-head competition. Third, environment-specific conditions (e.g. moistness of the semi-solid agar) allowed a flare to be generated during the selection experiment. These conditions would not have been present during the competitive motility assays and thus would have obscured advantageous characteristics.

4.4.2 Mutations underlying evolution of improved motility

Although the exact effects of mutations remain to be discovered, we can speculate on possible effects of mutations in focal genes that were not discussed before (Chapter 3). These mutations include mutations in FliG outside the MotA-FliG interaction site, FliM, FliC and CheB. As FliG is part of the rotor of the BFM, which is involved in switching the direction of flagellar rotation, torque generation and flagellar assembly [37–39], the A171T mutation in strain 18 might be involved in several functions. However, as the mutation is located adjacent to a region that was indicated to be involved in rotation bias in *Salmonella enterica* [40], it might be that the mutation alters the switching frequency of the BFM. Less structural information exists about the remaining proteins in which mutations were identified. FliM is part of the rotor, similar to FliG, and is involved in binding of the chemotaxis response regulator CheY [41]. Therefore, mutation A161V in strain 18 might result in a change in switching frequency. Alternatively, it might cause a structural change in the rotor, causing the foreign stators to have a better fit around the rotor. Because CheB is involved in the chemotaxis signalling pathway by resetting the chemotaxis receptors to their unstimulated state [35], mutation V311L in strain 18 might be involved in changing chemotaxis sensitivity. Lastly, FliC is the protein that constitutes the flagellar filament [42]. Consequently, mutation L10F in strain 18 might have resulted in an adaptive change in the structure of the filament or its mechanical properties. It would be interesting to examine this in more detail in the future and discover whether this change was adaptive specifically in the background of mutations that had occurred previously in this lineage.

We proposed previously (Chapter 3) that big-benefit mutations occurred more frequently in stator genes and *fliG*, but that not all mutations in these genes provided big benefits. Big-benefit mutations were defined as mutations that resulted in an increase in relative population swimming speed of >50% of the difference between the highest speed possible in that strain and the initial speed. In this chapter, we used competitive motility instead of relative population swimming speed as the measure of fitness. Although we could not claim causality of the mutations of which we resolved the order, expressed in terms of the contribution of mutations to increases in motility in semi-solid agar, the results suggest that big-benefit mutations occurred in *motA* homologues. Mutations that possibly resulted in a big benefit were identified as Laft E87K and Laft L284R in strain 08 (Figure 4.6) and MotA E91K in strain 18 (Figure 4.8). However, additional mutations in genes not investigated in this chapter might be present in the examined strains. Evidence for the suggested idea that not all mutations in the stator genes provided a big benefit, was found in mutations Laft A134S in strain 08 (Figure 4.6) and MotA Q237K in strain 18 (Figure 4.7; many consecutive flares were generated after gaining this mutation). Both mutations did not result in a high competitive motility.

4.4.3 Increase of complexity

Comparative genomic studies suggest a facilitating role for gene duplication followed by mutations in the evolution of increased complexity in protein assemblies. An experimental example of this mechanism was provided before [43]. In that study, the authors used ancestral gene resurrection in V-type ATPases, to show that complexity increased by gene duplication and subsequent mutations in both alleles. These mutations made both alleles obligatory to the evolved protein-complex. Because the stator genes in our study were expressed from a low copy plasmid

[33], several alleles were present (analogous to several gene duplications). This opened up the possibility for mutations in one or more alleles that would result in the increase of complexity similar to Finnigan et al. [43].

We found heterozygosity in six instances (Figure 4.5 and Figure 4.7). In one instance (strain 8.1e.2a.3a.4a.5a), the strain evolved improved motility with retention of heterozygosity. In the remaining five instances heterozygosity was found in endpoints, which makes it possible that the mutant allele (assuming a positive effect on motility) did not outcompete the non-mutated allele yet by plasmid incompatibility [44]. However, in 17 unique instances we found homozygous mutations in the stator genes, even though not all of these mutations led to high competitive motility (e.g. LaFT A134S in cBFM strain 08). Furthermore, mutation LaFT L284R in cBFM strain 08 was also found to be homozygous in different strains. Overall, our results provide preliminary evidence for increased complexity of the BFM, by the necessity of different alleles of one stator protein. Whether these cases indeed represent longer-term heterozygosity that is maintained by natural selection should be investigated further by propagation of these genotypes

4.5 Conclusion

The foreign stators that we examined had severe incompatibilities with the BFM and possibly other host-cell systems. If evolution were incapable of resolving these, then this would restrict the subset of stators in the biosphere that could be co-opted by compositional evolution to those that are compatible immediately. On the other hand, if evolution could resolve the interaction conflicts, then it would allow co-option of some of the incompatible stators. In Chapter 2, we have termed the boundary between the set of protein components that can be used for compositional evolution and those that cannot be used the *compatibility horizon*. The degree to which evolution can resolve conflicts between prospective components and existing complexes and host systems sets the upper-limit of the creative power of compositional evolution. The results we described in this chapter provide snapshots of the evolutionary dynamics that can take place at this compatibility horizon for our model system.

Our results directly show that the evolutionary trajectories described in in the previous chapters involved accumulating, step-wise adaptive genetic and phenotypic changes. We observed cases in which multiple mutations within a lineage affected a single BFM component and cases where multiple BFM components were sequentially mutated. The order in which BFM loci were mutated within lineages sometimes differed between replicate trajectories (e.g. stator first vs. FlhG first in strain 18). The phenotypic changes that were associated with these genetic changes differed within and between replicate trajectories of cBFM strains.

Taken together, these findings show that, in our model that looks at pre-adapted incompatible components, the epistatic interactions were such that they allowed stators at the compatibility horizon to be integrated by beneficial mutations that were to some degree additive. These mutations were available both within a single component (i.e. stator) and in different BFM and other cell-system components. The data also provide evidence of how epistatic interactions associated with some mutations can limit the set of components that can facilitate future adaptive change by mutation.

By definition, compositional evolution of protein complexes in nature must have involved pre-adapted components—if they were not, to some degree, pre-adapted, they would not be co-optable.

Our experiment used incompatible components (stators) derived from homologous complexes with the same function (BFMs of other species). Given that we know that stators have been acquired by horizontal gene transfer in some cases [45], our results suggest that the range over which this is possible is enlarged by the availability of parallel multi-step trajectories. By extension, our findings support the idea that, in general, the fraction of proteins in a biosphere that can be co-opted in compositional protein-complex evolution is likely to be enlarged by the availability of similar parallel multi-step mutational trajectories that push back the compatibility horizon.

4.6 Materials and Methods

4.6.1 Strains and culture conditions

The construction of the ancestral strains used in this chapter was described in more detail before (Chapter 2) and the strains are listed in Table 4.1. Briefly, the stator genes of *Escherichia coli* K12 MG1655 were replaced with stator genes from other bacteria. First, the original stator genes on the genome of *E. coli* K12 MG1655 were replaced by homologous recombination by a *rpsL-Neo* cassette [46], after which *rpsL* was replaced by *rpsL150*. Next, the foreign stator genes were expressed from the low-copy plasmid pBAD33 [33], which also harboured a chloramphenicol-resistance gene. The gene encoding the far-red fluorescent protein mKate2 was expressed from a plasmid that also harboured the stator genes of *E. coli* K12 MG1655. During the selection experiment (Chapter 2), the ancestral strains evolved (improved) motility. The strains that were isolated during the experiment are not listed in Table 4.1, because they are in essence similar to the ancestral strains.

All strains used in this chapter, both the strains listed in Table 2.3 and the strains that evolved from these ancestors during the selection experiment, were always grown overnight before use (O/N; 5 µl from -80°C stock; 16-17h; 250rpm; 37°C; orbit diameter 2.5cm) in 5 ml liquid LB medium (20 g/l LB powder) supplemented with 0.0025% chloramphenicol (w/v), unless stated otherwise.

Table 4.1. Strains used in this chapter.

Strain ID	Host strain	Stator genes on plasmid pBAD33	Donor strain stator genes
Strain 01-mKate2	<i>Escherichia coli</i> K12 str. MG1655 Δ <i>motAB</i> (short Δ <i>motAB</i>)	<i>motAB</i>	<i>Escherichia coli</i> K12 str. MG1655
Strain 08	Δ <i>motAB</i>	<i>lafTU</i>	<i>Escherichia coli</i> O111:H-str. 11128
Strain 11	Δ <i>motAB</i>	<i>motPS</i>	<i>Bacillus pseudofirmus</i> OF4
Strain 17	Δ <i>motAB</i>	<i>motAB</i>	<i>Listeria monocytogenes</i> EGD-e
Strain 18	Δ <i>motAB</i>	<i>motPS</i>	<i>Bacillus megaterium</i> DSM319

4.6.2 Competitive motility

Competitive motility measurements were described in detail in Chapter 2. Both a focal strain and a reference strain (strain01-mKate2) were grown O/N, after which they were diluted in liquid LB medium to an optical density at 600 nm (OD₆₀₀) of 1.0. Next, 5ul of both dilutions was inoculated 1 cm apart in the centre of a semi-solid agar plate (90 mm diameter petri dish). Semi-solid agar plates consisted of 27 ml of 50% LB medium (v/v; Tryptone, 10 g/l; Yeast Extract, 2.5 g/l; NaCl, 10 g/l), 0.3% agar (w/v), 0.2% L-arabinose (w/v) and 0.0025% chloramphenicol (w/v). The semi-solid agar plates were poured 19-20h in advance and dried statically at room temperature in stacks of 3 plates. After inoculation, the semi-solid agar plates were incubated statically for >24 hours in stacks of the same 3 plates in airtight, plastic boxes (DBP, 500 ml) at 37°C. To determine competitive motility of the focal strain, we used a Typhoon laser scanner (GE Healthcare Life Sciences) to scan the semi-solid agar plates. Next, a circle was fit to the interaction boundary between the competing strains on one plate. From the radius of this circle, competitive motility was calculated [47].

4.6.3 Single-cell measurements

Single-cell measurement were described in detail in Chapter 2. Briefly, a strain to be measured at the single-cell level was first grown O/N. Next, a sample of the culture was diluted 10x in liquid LB medium supplemented with 0.0025% chloramphenicol (w/v) and 0.2% L-arabinose (w/v). This culture was incubated for one hour at 37°C (250rpm, orbit diameter 2.5cm), after which a sample was washed twice by centrifugation and resuspension in liquid LB medium supplemented with 0.0025% chloramphenicol (w/v) and 0.2% L-arabinose (w/v). Optical density at 600 nm of this sample was adjusted to a value of 0.2-0.3, using the same medium, after which the sample was put on ice. Next, this sample was added to a BSA-coated channel that was prepared one day in advance and measurements of single-cell phenotypes were performed immediately after filling the BSA-coated channel, with a temperature-controlled (37°C) widefield microscope (OKO) using a 20x objective. Each measurement was performed for 30 seconds at 8 frames per second (fps) and was saved as a TIFF stack. The recorded TIFF-stack was analysed by ImageJ MOSAIC ParticleTracker and the obtained data was loaded into a self-developed MATLAB script for further analysis.

4.6.4 Single-BFM measurements

Strains to be measured were grown O/N, after which a sample was diluted 1000x in 5 ml liquid LB medium supplemented with 0.2% L-arabinose (w/v). This dilution was incubated for 1h (250rpm; 37°C; orbit diameter 2.5cm) in order to express the stator genes. Next, flagella were sheared by pipetting 1 ml of the culture 50x up-and-down through a 0.6x25 mm needle. Cells with sheared flagella were washed twice by centrifugation and resuspension in motility buffer of pH 7 (10mM K₂HPO₄, 85 mM NaCl, 0.1 mM EDTA). A sample was pipetted onto a microscope slide and covered with a glass coverslip that rested on two small glass coverslips (Figure 4.10). Measurements were performed under a light microscope at room temperature with a 100x objective. We actively searched for cells that were stuck to the microscope slide with one sheared flagellum and therefore were spinning around this anchor point. Spinning cells were recorded for at least 15 seconds with a frame rate of 76 frames per second. Analysis of the frequency of rotation and tumbling events was performed with a custom-made Python script [48].

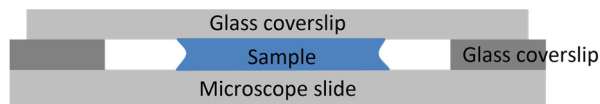


Figure 4.10. The construction used for single-BFM measurements. The sample containing cells with sheared flagella is covered with a glass coverslip, creating a ‘column’ of water.

4.6.5 High Resolution Melt Analysis

The procedure for performing High Resolution Melt Analysis (HRMA) was described elsewhere (Chapter 2). Briefly, O/N cultures of a reference strain, which is the ancestral strain (Table 4.1) of the focal strain, and a focal strain were added to the Solis BioDyne HOT FIREPol® EvaGreen® HRM Mix (No ROX), together with primers appropriate for the focal mutation. HRMA was performed on Illumina ECO real-time PCR system and per strain we performed three replicates. After melting of the PCR product, the melt curves were normalized automatically to obtain 100% fluorescence before melting and 0% fluorescence after melting. These normalized fluorescence curves were then used to calculate the Euclidian distance between a standard (we used the average of the reference strain) and all strains that were measured [49]. Next, we tested for significant differences in Euclidian distance between the reference strain and the focal strain. Together with visual inspection of the melt curve, mutations were identified.

4

4.6.6 Statistical Analysis

Different statistical analyses were performed, which depended on the distribution of the data. See text for details. When compensating for multiple comparisons, the Benjamini-Hochberg procedure was used [50].

4.7 Literature

- 1 Watson RA (2002) *Compositional Evolution: Interdisciplinary Investigations in Evolvability, Modularity and Symbiosis*.
- 2 Pallen M & Matzke N (2006) From The Origin of Species to the origin of bacterial flagella. *Nat. Rev. Microbiol.* **4**, 784–790.
- 3 Liu R & Ochman H (2007) Stepwise formation of the bacterial flagellar system. *Proc. Natl. Acad. Sci.* **104**, 7116–7121.
- 4 Mulkidjanian A & Makarova K (2007) Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. *Nat. Rev. Microbiol.* **5**, 892–899.
- 5 Dolezal P, Likic V, Tachezy J & Lithgow T (2006) Evolution of the molecular machines for protein import into mitochondria. *Science* **313**, 314–8.
- 6 Clements A, Bursac D, Gatsos X, Perry AJ, Civciristov S, Celik N, Likic VA, Poggio S, Jacobs-Wagner C, Strugnell RA & Lithgow T (2009) The reducible complexity of a mitochondrial molecular machine. *Proc. Natl. Acad. Sci.* **106**, 15791–15795.

- 7 Archibald JM, Logsdon Jr. JM & Doolittle WF (2000) Origin and Evolution of Eukaryotic Chaperonins: Phylogenetic Evidence for Ancient Duplications in CCT Genes. *Mol. Biol. Evol.* **17**, 1456–1466.
- 8 Gabaldón T, Rainey D & Huynen M (2005) Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I). *J. Mol. Biol.* **348**, 857–70.
- 9 Seidl MF & Schultz J (2009) Evolutionary flexibility of protein complexes. *BMC Evol. Biol.* **9**, 155.
- 10 Blair DF (2003) Flagellar movement driven by proton translocation. *FEBS Lett.* **545**, 86–95.
- 11 Chun S & Parkinson J (1988) Bacterial motility: membrane topology of the Escherichia coli MotB protein. *Science* **239**, 276–278.
- 12 Khan S, Dapice M & Reese TS (1988) Effects of mot Gene Expression on the Structure of the Flagellar Motor at the Marine Biological Laboratory. *J. Mol. Biol.* **202**, 575–584.
- 13 de Mot R & Vanderleyden J (1994) The C-terminal sequence conservation between OmpA-related outer membrane proteins and MotB suggests a common function in both Gram-positive and Gram-negative bacteria, possibly in the interaction of these domains with peptidoglycan. *Mol. Microbiol.* **12**, 333–334.
- 14 Koebnik R (1995) Proposal for a peptidoglycan-associating alpha-helical motif in the C-terminal regions of some bacterial cell-surface proteins. *Mol. Microbiol.* **16**, 1269–1270.
- 15 Zhou J, Sharp LL, Tang HL, Lloyd SA, Billings S, Braun TF & Blair DF (1998) Function of Protonatable Residues in the Flagellar Motor of Escherichia coli: a Critical Role for Asp 32 of MotB. *J. Bacteriol.* **180**, 2729–2735.
- 16 Zhou J & Blair DF (1997) Residues of the Cytoplasmic Domain of MotA Essential for Torque Generation in the Bacterial Flagellar Motor. *J. Mol. Biol.* **273**, 428–439.
- 17 Lloyd SA & Blair DF (1997) Charged Residues of the Rotor Protein FlgG Essential for Torque Generation in the Flagellar Motor of Escherichia coli. *J. Mol. Biol.* **266**, 733–744.
- 18 Zhou J, Lloyd SA & Blair DF (1998) Electrostatic interactions between rotor and stator in the bacterial flagellar motor. *Proc. Natl. Acad. Sci.* **95**, 6436–6441.
- 19 Adler J (1966) Chemotaxis in bacteria. *Science* **153**, 708–716.
- 20 Adler J (1966) Effect of amino acids and oxygen on chemotaxis in Escherichia coli. *J. Bacteriol.* **92**, 121–129.
- 21 Berg HC & Brown DA (1972) Chemotaxis in Escherichia coli analysed by three-dimensional tracking. *Nature* **239**, 500–504.
- 22 Park JY, Yoo SJ, Hwang CM & Lee S-H (2009) Simultaneous generation of chemical concentration and mechanical shear stress gradients using microfluidic osmotic flow comparable to interstitial flow. *Lab Chip* **9**, 2194–202.
- 23 Righetti P, Brost B & Snyder R (1981) On the limiting pore size of hydrophilic gels for electrophoresis and isoelectric focusing. *J. Biochem. Biophys. Methods* **4**, 347–363.
- 24 Pernodet N, Tinland B, Sadron IC & Pasteur CL (1997) Pore size of agarose gels by atomic force microscopy. *Electrophoresis* **18**, 55–58.
- 25 Narayanan J, Xiong J-Y & Liu X-Y (2006) Determination of agarose gel pore size: Absorbance measurements vis a vis other techniques. *J. Phys. Conf. Ser.* **28**, 83–86.
- 26 Croze OA, Ferguson GP, Cates ME & Poon WCK (2011) Migration of chemotactic bacteria in soft agar: role of gel concentration. *Biophys. J.* **101**, 525–34.

- 27 Wadhams GH & Armitage JP (2004) Making sense of it all: bacterial chemotaxis. *Nat. Rev. Mol. Cell Biol.* **5**, 1024–37.
- 28 Adler J (1975) Chemotaxis in bacteria. *Annu. Rev. Biochem.* **44**, 341–356.
- 29 Wolfe AJ & Berg HC (1989) Migration of bacteria in semisolid agar. *Proc. Natl. Acad. Sci.* **86**, 6973–6977.
- 30 Braun TF, Poulson S, Gully JB, Empey JC, Van Way S, Putnam a & Blair DF (1999) Function of proline residues of MotA in torque generation by the flagellar motor of *Escherichia coli*. *J. Bacteriol.* **181**, 3542–51.
- 31 Brown PN, Terrazas M, Paul K & Blair DF (2007) Mutational analysis of the flagellar protein FliG: sites of interaction with FliM and implications for organization of the switch complex. *J. Bacteriol.* **189**, 305–12.
- 32 Wittwer CT (2009) High-resolution DNA melting analysis: advancements and limitations. *Hum. Mutat.* **30**, 857–9.
- 33 Guzman LM, Belin D, Carson MJ, Beckwith J, Guzman L, Belin D & Carson MJ (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130.
- 34 Sowa Y & Berry RM (2008) Bacterial flagellar motor. *Q. Rev. Biophys.* **41**, 103–32.
- 35 Tindall MJ, Porter SL, Maini PK, Gaglia G & Armitage JP (2008) Overview of mathematical approaches used to model bacterial chemotaxis I: the single cell. *Bull. Math. Biol.* **70**, 1525–69.
- 36 de Visser JAGM & Lenski RE (2002) Long-term experimental evolution in *Escherichia coli*. XI. Rejection of non-transitive interactions as cause of declining rate of adaptation. *BMC Evol. Biol.* **2**, 19.
- 37 Yamaguchi S, Aizawa S, Kihara MAY, Isomura M, Jones CJ & Macnab RM (1986) Genetic Evidence for a Switching and Energy-Transducing Complex in the Flagellar Motor of *Salmonella typhimurium*. *J. Bacteriol.* **168**, 1172–1179.
- 38 Irikura VM, Kihara M, Yamaguchi S, Sockett H & Macnab RM (1993) *Salmonella typhimurium* fliG and fliN Mutations Causing Defects in Assembly, Rotation, and Switching of the Flagellar Motor. *J. Bacteriol.* **175**, 802–810.
- 39 Sockett H, Yamaguchi S, Kihara MAY, Irikura VM & Macnab RM (1992) Molecular Analysis of the Flagellar Switch Protein FliM of *Salmonella typhimurium*. *J. Bacteriol.* **174**, 793–806.
- 40 Togashi F, Yamaguchi S, Kihara MAY & Aizawa S (1997) An extreme clockwise switch bias mutation in fliG of *Salmonella typhimurium* and its suppression by slow-motile mutations in motA and motB. *J. Bacteriol.* **179**, 2994–3003.
- 41 Bren A & Eisenbach M (1998) The N terminus of the flagellar switch protein, FliM, is the binding domain for the chemotactic response regulator, CheY. *J. Mol. Biol.* **278**, 507–14.
- 42 Berg HC (2003) The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19–54.
- 43 Finnigan GC, Hanson-Smith V, Stevens TH & Thornton JW (2012) Evolution of increased complexity in a molecular machine. *Nature* **481**, 360–4.
- 44 Velappan N, Sblattero D, Chasteen L, Pavlik P & Bradbury RM (2007) Plasmid incompatibility: more compatible than previously thought? *Protein Eng. Des. Sel.* **20**, 309–313.
- 45 Paulick A, Koerdt A, Lassak J, Huntley S, Wilms I, Narberhaus F & Thormann KM (2009) Two different stator systems drive a single polar flagellum in *Shewanella oneidensis* MR-1. *Mol. Microbiol.* **71**, 836–50.
- 46 Heermann R, Zeppenfeld T & Jung K (2008) Simple generation of site-directed point mutations in the *Escherichia coli* chromosome using Red(R)/ET(R) Recombination. *Microb. Cell Fact.* **7**, 14.
- 47 Korolev K, Müller M, Karahan N, Murray A, Hallatschek O & Nelson D (2012) Selective sweeps in growing microbial

colonies. *Phys. Biol.* **9**, 26008.

48 Luik S (2014) Torque and tumbling in evolved chimeric bacterial flagellar motors. .

49 Hjelmsø MH, Hansen LH, Bælum J, Feld L, Holben WE & Jacobsen CS (2014) High Resolution Melt analysis for rapid comparison of bacterial community composition. *Appl. Environ. Microbiol.* **80**, 3568–3575.

50 Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300.

5 Diversity of first-step adaptive mutations in foreign components during compositional evolution

Régis C.E. Flohr, Min-Young Heo, Erwin van Rijn, Francesco Pedaci, Hubertus J.E. Beaumont. *In preparation for publication.*

5.1 Abstract

In previous chapters of this thesis, we provided experimental evidence for the potential and creativity of compositional evolution. Using our model system, the bacterial flagellar motor (BFM), we showed that integration of incompatible foreign components could occur along multiple genetic routes, even for individual foreign components. However, we did not resolve the direct contribution of individual mutations to the integration of the foreign components. In this chapter, we focussed on mutations in the foreign component during the first adaptive step. We demonstrate that these first-step mutations differ for different foreign components but also differed between replicate integration events of the same component. Furthermore, we show that different mutations had different phenotypic effects and that mutations were not restricted to a specific part of the foreign components. The findings shed light on the degree of diversification between homologous components in orthologous protein complexes with the same function.

5.2 Introduction

Comparative genomic studies suggest that protein complexes, such as the bacterial flagellar motor (BFM) [1] or V-type ATPase [2,3], evolved by the stepwise addition of pre-existing protein components to simpler assemblies [4–11], a process called compositional evolution [12]. However, experimental evidence of the mechanisms underlying protein-complex evolution by compositional evolution is scarce (e.g. [3,13–16]) and additional experimental insight into its details is necessary for a more complete understanding.

In the previous chapters of this thesis we introduced and exploited an experimental system, with which we aimed to generate additional experimental insight. First, we developed our experimental system, with a particular focus on the stator complex of the BFM of *E. coli*. The BFM is the main organelle for bacterial motility and consists of more than 20 different protein components [17]. The stator complex of the BFM is composed of two proteins, MotA and MotB [18,19], and has two major functions. It is involved in anchoring of the BFM to the cell wall [20,21] and in coupling of the translocation of ions from outside to inside the cell to flagellar rotation and torque generation [17,22–24]. As such, the stator complex is essential for BFM-mediated motility. In chapter 2, we replaced the stator genes of *E. coli* by stator genes from different species across the bacterial kingdom. This resulted in strains with a chimeric BFM (cBFM strains). For these cBFM strains, we investigated the possibility of immediate compatibility of foreign stators (genes and/or proteins), determined by BFM-mediated motility of populations of cells in semi-solid agar. We also investigated the capacity of these chimeric systems to evolve compatibility and/or improved motility, which we referred to as *evolutionary integration*. During the selection experiment, we selected for cells that evolved an improved ability to spread out in semi-solid agar compared with their direct ancestor and thereby formed *flares*. Flares were defined as wedge-shaped populations of cells capable of escaping from the chemotactic front of their ancestral population in semi-solid agar. Genotypes isolated from these flares were used to start additional rounds of selection, which in some cases yielded additional flares. A sequence of genotypes isolated from consecutive flares (primary, secondary, etc.) was called an *evolutionary lineage* and the final genotype in an evolutionary lineage was referred to as *endpoint*. By definition, an endpoint was not able to evolve improved motility in our selection experiment (for details of the experiment see Chapter 2). Note that this does mean that further adaptive evolution was no longer possible, but rather that the number of accessible beneficial mutations, and the size of their beneficial effects, had decreased dramatically. The evolved endpoints were analysed with whole-genome re-sequencing in order to identify mutations that had occurred (Chapter 3) and for a number of evolutionary lineages, the order in which mutations had occurred in genes involved in the BFM or chemotaxis signalling pathway was resolved (Chapter 4). For a subset of the lineages for which the mutations had been ordered, we quantified the accompanying competitive motility against a fluorescently-labelled reference strain that harboured wildtype *E. coli* stators. This allowed us to gain insight into the phenotypic changes that correlated with these mutations. However, this did not reveal directly the phenotypic effect of individual—mutations at loci other than the focal genes may also have played a role in increasing the cellular motility.

In this chapter, we investigate directly the contribution of a subset of the mutations that facilitated the first step in functional integration of the foreign stators. We focussed on genotypes isolated from primary flares that evolved improved motility by mutations in the genes encoding the

foreign stators. These mutations either resulted in compatibility, in cases where the foreign stators were incompatible, or improved motility when the stators were compatible immediately. We were interested in the diversity of causal mutations by which sub-compatible foreign stators could be integrated. Are there multiple causal mutations accessible in a single foreign component? Do the causal mutations in different foreign stators overlap and which structural/functional domains of the stators are involved? What is the relative importance of point mutations *versus* insertions and deletions? We also examined the effect of these mutations on chemotactic motility.

Our results showed that multiple causal mutations in both MotA and MotB homologues, sometimes at the same position in the protein, evolved within a single cBFM strain, which indicated that there were several possibilities to functionally integrate a foreign component. However, because in some instances identical mutations were discovered, evolution was shown to be repeatable as well. Strain-specific constraints were identified by the diversity in causal mutations between different cBFM strains. The majority of causal mutations we discovered were identified as single nucleotide polymorphisms (SNPs) and we showed that different mutations resulted in different motility levels. Finally our results revealed that the mutations were not restricted to a specific part of the foreign stators.

5.3 Results

In order to identify mutations in the stator genes that caused emergence of the first flare, we determined whether introduction of stator expression vectors isolated from primary flares, conferred improved motility in the unevolved host *E. coli* background. We did this by comparing the motility of the resulting strains (*i.e.* distance travelled in 24h in semi-solid agar plates) with that of the unevolved host harbouring the unevolved stator expression vector. When introduction of the evolved vector caused increased motility, mutations had to be present on the plasmid as this was the only locus at which genetic differences could exist. These mutations either could result in compatibility (when the ancestral strain was non-motile initially) or improved motility (when the ancestral strain was motile immediately). Next, we examined the plasmid sequence data in the NGS analysis described in Chapter 3. This analysis included the evolved plasmids present in the end point genotypes. This allowed us to identify mutations in the stator genes and, possibly, in other parts of the expression vector. Mutations in the stator genes were verified by Sanger sequencing.

The causal stator mutations identified in this manner are shown in Figure 5.1. With the exception of an insertion and deletions in cBFM strain 17, all mutations identified by Sanger sequencing had already been detected by whole-genome re-sequencing (Chapter 3). The presence of additional mutations outside of the stator genes on the plasmids was excluded by investigating the whole-genome re-sequencing dataset mentioned above and by performing additional experiments (in case of strain 29).

We will start by discussing the cBFM strains that harboured compatible stators and were thus motile directly (compatibility group B; see Chapter 2). Next, we will turn to the strains that contained incompatible stators and were non-motile initially (compatibility group C; see Chapter 2).

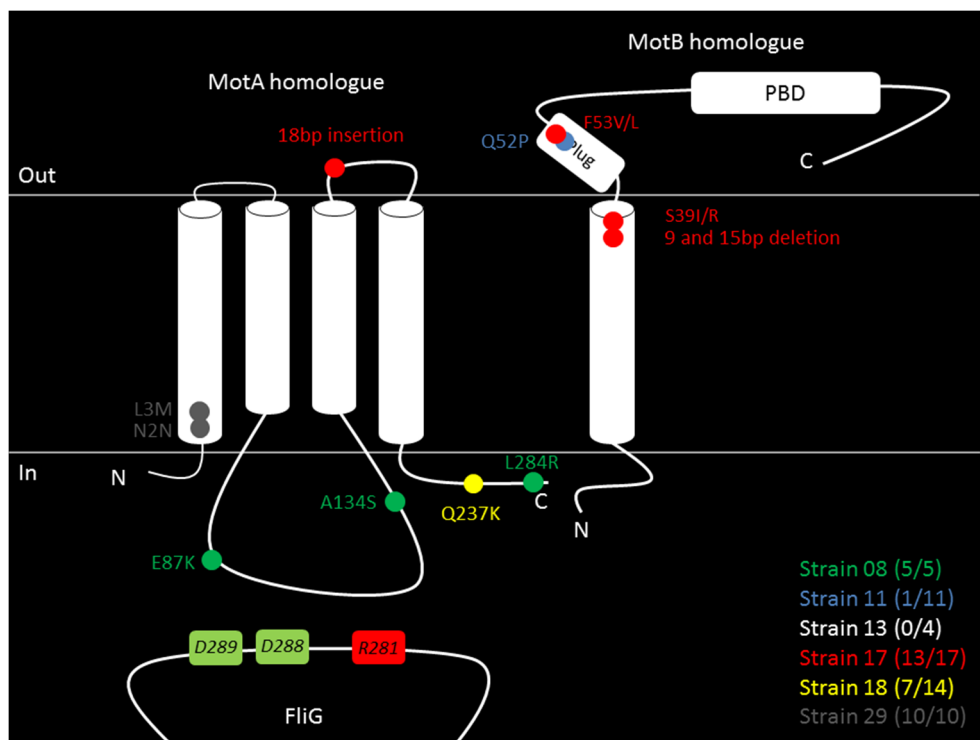


Figure 5.1. Location of causal mutations in the stator proteins of strains that evolved compatibility or improved motility. Horizontal lines indicate the cell membrane that separates the cytoplasm (in) from the periplasmic space (out). Strain x (y/z) indicates for strain x, the number of the genotypes isolated from primary flares that contained causal mutations in the stator proteins (y). The total number of genotypes isolated from primary flares, that were more motile than their direct ancestor, is indicated by z. The structure was obtained by predicting the transmembrane (TM) helices (indicated by cylinders) through the online TMHMM Server v.2.0 [25]. By aligning the foreign stator sequences to the stator sequences of *E. coli*, we assigned amino acid positions to the mutations in the foreign stators. Because gaps were involved in all alignments, the actual positions are probably different. However, the regions of the proteins in which mutations occurred (e.g. TM helix or linker-domain between TM helices) are likely to be accurate. The residue in red (R281) is positively charged, while the residues in green (D288 and D289) are negatively charged. PDB: peptidoglycan binding domain. Plug: plug domain. Positions of mutations, plug and PBD were determined by alignment to *E. coli* stators. See text for details.

5.3.1 Compatibility group B: strain 11

From genotypes isolated from primary flares of strain 11 (e.g. strain 11.1i) plasmids were isolated and transferred to *E. coli* cells lacking stator genes. The nomenclature used to refer to the resulting genotypes is based on the identifier of the first-flare genotype (e.g. 11.1i) to which the suffix “-P” is added (e.g. 11.1i-P).

Of the 11 genotypes isolated from primary flares that emerged from ancestral cBFM strain 11, only one (9% of total) harboured causal mutations in the stator genes (Table 5.1). The mutation Q52P was discovered in MotB-homologue MotS and likely affects the plug domain of MotS (Figure 5.1). The plug domain is involved in preventing the flow of ions through the stator complex when not

bound to the BFM and might also affect the interaction of the peptidoglycan-binding domain (PBD) of MotS with the peptidoglycan layer upon docking to the BFM [26]. This interaction, on its turn, might affect positioning of the stator around the rotor [27].

Next, we investigated phenotypic effects of the mutation by measuring the population swimming speed of cells harbouring the mutation during the first 8h after inoculation. The population swimming speed was determined by measuring the distance the front of a population of cells had travelled per hour in semi-solid agar (*Materials and Methods*). In case of strain 11.1i-P, the population showed *normal* motility. This type of motility is characterised by a ring of cells in front of the rest of the population and allowed the population swimming speed to be measured. The population-swimming-speed of strain 11.1i-P, relative to the population swimming speed of a strain harbouring *E. coli* wildtype stators, was 0.55 ± 0.032 ($n = 9$; mean \pm standard deviation). This was substantially faster than strain 11, which was either immotile or performed atypical motility (ATM). ATM is characterized by a dotted appearance (3D bacterial colonies surrounding the inoculation point) in semi-solid agar after 24h and was described in literature to indicate that a very small fraction of the inoculated cells was motile and that motile cells usually produced daughter cells that were non-motile [23,28,29]. Other explanations for ATM might be low tumbling frequency of cells and low torque generated by the BFM (Chapter 2). The population swimming speed of ATM-performing populations was too slow to be measured with our method (*Materials and Methods*).

Table 5.1. Causal mutation in the stator genes of strain 11, with an indication of motility as a result of the mutation. ATM: Atypical motility. Motility: normal > ATM (see text for details). Strain 11 performed ATM only stochastically.

Strain	Mutation		Motility
	MotA homologue	MotB homologue	
11			ATM (44%)
11.1i-P		Q52P	normal

5.3.2 Compatibility group B: strain 17

In contrast to cBFM strain 11, 87% of the primary flares (13 out of 15) of strain 17 were found to contain genotypes harbouring causal mutations in the stator genes (Table 5.2). Four strains harboured the MotB F53V substitution and two strains harboured a different variant, MotB F53L. In one of these latter strains (17.1h-P) the mutation was heterozygous, which indicated that this strain contained MotB genes coding for proteins with the original amino acid (F), but also MotB genes coding for proteins with the mutated amino acid (V). Similar to strain 11, the mutation at position 53 of MotB was located in the plug domain (Figure 5.1). Therefore, it might affect the flow of ions through the stator complex when not bound to a BFM or affect the interaction between the PBD and the peptidoglycan layer upon binding to the BFM, resulting in a changed positioning of the stator around the rotor [26,27]. At amino acid position 39 of MotB we also found two different mutations, S39I (both homozygous and heterozygous with original amino acid) and S39R. Transmembrane-domain prediction indicated that this position was located to the C-terminal end of the transmembrane helix (Figure 5.1). The remaining four causal mutations had not been discovered before by whole-genome re-sequencing (Chapter 3). Two of the mutations involved an identical, out-

of-frame 18bp-tandem-duplication (ACTCATTCATGCGATGGG) located to the second periplasmic loop of MotA (Figure 5.1), effectively inserting 6 amino acids (AMGLIHAMGE → AMGLIHAMGLIHAMGE; duplicated amino acids are underlined). The other two mutations we identified here were unique, but did have an overlap. One mutation involved a 9bp deletion (CCTCCAGTT) located near the end of the transmembrane domain of MotB (Figure 5.1), effectively deleting 3 amino acids (FASSSVD → FAVD; deleted amino acids are underlined). The other mutation involved a 15bp deletion (AGTTCAGTTGATGCA; the underlined bases indicate overlap with the 9bp deletion) that started at residue 39 in the transmembrane domain and resulted in a deletion of 5 amino acids (SSSVDAK → SK; deleted amino acids are underlined). In both cases the distance between transmembrane domain and plug domain was shortened by the deletion.

Table 5.2. Causal mutations in the stator genes of strain 17, with an indication of motility as a result of the mutation. ATM: Atypical motility. Motility: normal > slow > ATM (see text for details). Heterozygous: both original and mutated amino acid were present. 18bp insertion started within the codon coding for glycine at position 172. 9bp deletion started within the codon coding for alanine at position 37. 15bp deletion started at residue 39.

Strain	Mutation		Motility
	MotA homologue	MotB homologue	
17			ATM
17.1a-P		S39I (heterozygous)	normal
17.1b-P		F53V	normal
17.1c-P		F53V	normal
17.1e-P		F53V	normal
17.1f-P		S39R	normal
17.1g-P	18bp insertion (Glycine 172)		slow
17.1h-P		F53L (heterozygous)	normal
17.1i-P		F53V	normal
17.1j-P	18bp insertion (Glycine 172)		slow
17.1k-P		9bp deletion (Alanine 37)	normal
17.1l-P		F53L	normal
17.1n-P		S39I	normal
17.1o-P		15bp deletion (Serine 39)	normal

In order to investigate the phenotypic effects of these causal mutations, we measured the relative population-swimming-speeds in semi-solid agar (

Figure 5.2). We first tested the hypothesis that different mutations resulted in different relative population-swimming-speeds. When we performed the overall analysis, only including strains of which the population speeds could be measured, we indeed found different phenotypic effects caused by different mutations (one-way ANOVA; $n = 3$; $p < 0.001$). Strains 17.1g-P and 17.1j-P, both harbouring the 18bp insertion in MotA, were too slow to be measured according to our protocol (but moved further away from the inoculation point than ATM-performing populations after 24h) and were thus assumed to be different phenotypically from the other genotypes. As expected we found that identical mutations resulted in indistinguishable phenotypes (Benjamini-Hochberg-corrected Student's t -tests; $n = 12$; $p > 0.05$ for all comparisons). This result was also obtained for the F53V and

F53L substitutions (Benjamini-Hochberg-corrected Student's *t*-tests; $n = 12$; $p > 0.05$ for all comparisons), which indicated that not all different mutations resulted in different phenotypes. In contrast to the previous result, in which one amino acid was mutated twice, we did find significant differences in relative population-swimming-speeds caused by either the S39I or S39R mutation. The latter mutation changed the charge of the amino acid at this position and yielded a lower speed (Benjamini-Hochberg-corrected Student's *t*-tests; $n = 12$; $p < 0.001$). Additionally, strains with heterozygous genotypes (17.1a-P and 17.1h-P) were significantly slower than the strains with homozygous genotypes (17.1n-P and 17.1l-P, respectively; Benjamini-Hochberg-corrected Student's *t*-tests; $n = 12$; $p < 0.001$ for both comparisons). The mutation that resulted in the highest population-swimming-speed was the 15bp deletion located to the transmembrane helix of MotB (Benjamini-Hochberg-corrected Student's *t*-tests; $n = 12$; $p < 0.001$ for all comparisons), with the 9bp deletion in this same region resulting in a slightly lower speed than the one caused by the 15bp deletion. These results are remarkable, as such large deletions often have deleterious effects [30,31] (but see also [32], although the proposed mechanism is not a structural change).

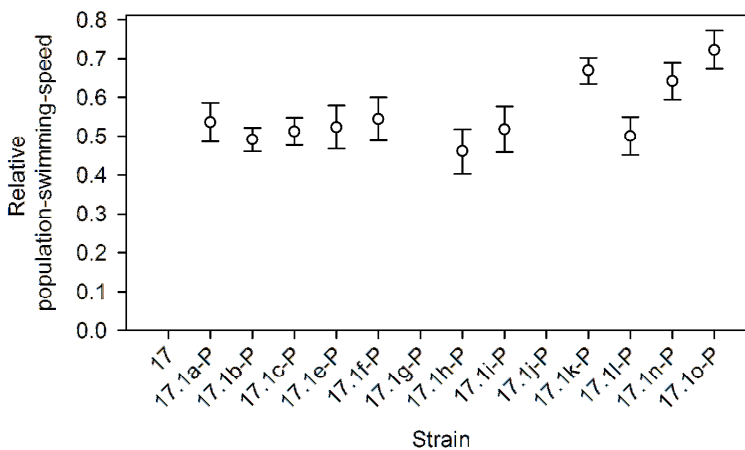


Figure 5.2. Population swimming speed, relative to a strain harbouring wildtype *E. coli* stators, as a result of causal mutations in the stator genes of strain 17. Strains 17, 17.1g-P and 17.1j-P were too slow to be measured according to our protocol. Bars indicate standard deviation from mean. $n = 12$. See text for details.

5.3.3 Compatibility group B: strain 18

Fourteen primary flares were generated in strain 18, of which 7 (50%) contained genotypes harbouring causal mutations in the foreign stator genes (Table 5.3). In all strains, only one causal mutation was found, adding a positive charge to a position that was uncharged before. This mutation, Q237K, resided in the C-terminal region of MotA (Figure 5.1), a region that was indicated to be important for motility in chimeric motors using PomA from *Vibrio alginolyticus* and FliG from *E. coli* [33]. In that study, a charge reversing mutation at position 232 (R232E; positive to negative) resulted in a much-reduced motility.

Accompanying the mutation in MotA, we found a relative population-swimming-speed of strains carrying the mutation that was too slow to be measured. However, the strains were substantially faster than the ATM-performing ancestral strain 18 (observation).

Table 5.3. Causal mutations in the stator genes of strain 18, with an indication of motility as a result of the mutation. ATM: Atypical motility. Motility: slow > ATM (see text for details).

Strain	Mutation		Motility
	MotA homologue	MotB homologue	
18			ATM
18.1e-P	Q237K		slow
18.1f-P	Q237K		slow
18.1g-P	Q237K		slow
18.1i-P	Q237K		slow
18.1j-P	Q237K		slow
18.1k-P	Q237K		slow
18.1m-P	Q237K		slow

In the next section we investigated strains from compatibility group C, in order to examine the causal mutations in strains that harboured incompatible stators and were thus non-motile initially.

5.3.4 Compatibility group C: strain 08

cBFM strain 08 generated the lowest number of primary flares in compatibility group C, partially due to the fact that only 3 of the 10 replicate populations evolved compatibility in semi-solid agar. However, all genotypes isolated from the primary flares harboured causal mutations in the foreign stator genes (Table 5.4).

Three different mutations were discovered in MotA-homologue LafT, of which substitution A134S was present in all five strains. In two strains it was present in isolation (8.1a-P and 8.1e-P), which allowed us to determine phenotypic effects of this substitution. Strains harbouring the LafT A134S mutation all performed ATM, indicating a small but positive effect on population swimming speed, as ancestral strain 08 is non-motile. The mutation is located in the cytoplasmic loop of LafT, in between the LafT-FliG interaction region and the third transmembrane helix (Figure 5.1). Previously [34], mutations in this region were found to suppress the negative effects of mutations in the C-terminal domain of MotA, which was suggested to be due to the proper alignment of the MotA-FliG interaction region. In three different strains, we found two mutations accompanying mutation A134S, both of which were located in LafT. Two strains harboured the L284R mutation, a mutation at the C-terminus of LafT that was previously indicated to be important in motility (see above) [34]. A third strain harboured mutation E87K, resulting in a charge reversal (negative to positive) near the LafT-FliG interaction site.

Table 5.4. Causal mutations in the stator genes of strain 08, with an indication of motility as a result of the mutation. Strain 08 was non-motile. ATM: Atypical motility. Motility: normal > ATM > non-motile (see text for details).

Strain	Mutation		Motility
	MotA homologue	MotB homologue	
8			Non-motile
8.1a-P	A134S		ATM
8.1b-P	A134S / L284R		normal
8.1c-P	A134S / L284R		normal
8.1d-P	A134S / E87K		normal
8.1e-P	A134S		ATM

As indicated in Table 5.4, the mutations accompanying mutation A134S resulted in normal motility in semi-solid agar, thereby representing an improvement compared with ATM. While substitution E87K resulted in a higher motility compared with substitution L284R (Figure 5.3; Benjamini-Hochberg-corrected Student's *t*-tests; $n = 9$; $p < 0.001$ for both comparisons), we unexpectedly also found a significant difference between strains 8.1b-P and 8.1c-P (Benjamini-Hochberg-corrected Student's *t*-tests; $n = 9$; $p = 0.014$). As the mutations found in these strains were identical, a possible explanation for this discrepancy might be additional mutations that were gained during the transformation of cells lacking stator genes with evolved plasmids. However, results of the relative population-swimming-speed of strains 8.1b and 8.1c (donor strains of the plasmids in strains 8.1b-P and 8.1c-P, respectively; Supplementary information Figure 5.4; $n = 3$) suggested a small difference as well. Another explanation for the discovered discrepancy might be epigenetic changes on the plasmids. Epigenetic changes (e.g. DNA methylation) can alter gene expression and can be inherited from generation to generation [35,36], but cannot be detected by sequencing as it does not change the sequence of bases in the DNA.

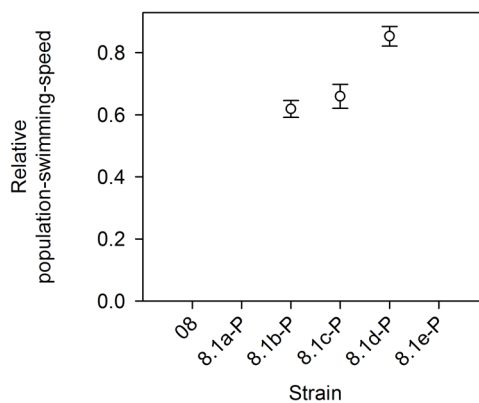


Figure 5.3. Population swimming speed, relative to a strain harbouring wildtype *E. coli* stators, as a result of causal mutations in the stator genes of strain 08. Strain 08 was non-motile. Strains 8.1a-P and 8.1e-P were too slow to be measured according to our protocol. Bars indicate standard deviation from mean. $n = 9$. See text for details.

5.3.5 Compatibility group C: strain 29

cBFM strain 13 generated four primary flares that contained genotypes that were more motile than the ancestral strain. However, not one of these genotypes harboured causal mutations in the foreign stators. The situation was different for strain 29, in which 10 out of 10 primary flares contained genotypes that harboured causal mutations in or just upstream of the stator genes (Table 5.5). In contrast to the causal mutations we discussed so far, the causal mutations in strain 29 were located to the very beginning of MotA or even upstream of MotA, thus not affecting the coding sequence.

Five strains harboured the synonymous N2N mutation (AAC → AAT), which indicated that the amino acid was not changed by the base change. This mutation is not favoured by the codon bias in *E. coli*, as the ratio for encoding asparagine is 55% AAC and 45% AAT [37]. In two strains we discovered a causal mutation at the position adjacent to N2. This substitution, L3M, resulted in a change in the size of the amino acid, but not the charge. The third causal mutation, located 6bp upstream of the start codon GTG, might have affected the Shine-Dalgarno (SD) sequence of MotA and possibly affected MotA expression levels.

Additional mutations on the plasmid were discovered in the p15A origin of replication during whole-genome re-sequencing (Chapter 3). As these mutations might affect the copy number of plasmids in a cell (and thereby influence the number of stator genes), they may influence the expression level of stator proteins. This on its turn may result in compatibility of the foreign stators in this strain. However, no such mutations were found upon additional screening by Sanger sequencing.

The three causal mutations that we identified in strain 29 resulted in ATM, which indicated that the mutations resulted in compatibility of the foreign stators as strain 29 is non-motile. However, the mutation in the putative SD-sequence did not result in ATM in all instances ($n = 12$), revealing that this mutation caused motility only in a stochastic manner. A possible explanation for this phenotype is stochastic rotor-stator interaction.

Table 5.5. Causal mutations in the stator genes of strain 29, with an indication of motility as a result of the mutation. Strain 29 was non-motile. Mutations in the putative SD-sequence of MotA (6bp upstream of startcodon) cause motility stochastically. ATM: Atypical motility. Motility: ATM > non-motile (See text for details).

Strain	Mutation		Motility
	MotA homologue	MotB homologue	
29			Non-motile
29.1a-P	L3M		ATM
29.1b-P	L3M		ATM
29.1d-P	N2N		ATM
29.1e-P	N2N		ATM
29.1f-P	N2N		ATM
29.1g-P	N2N		ATM
29.1i-P	N2N		ATM
29.1k-P	6bp upstream of startcodon		ATM (33%)
29.1l-P	6bp upstream of startcodon		ATM (78%)
29.1m-P	6bp upstream of startcodon		ATM (56%)

5.4 Discussion

In this chapter we investigated causal mutations in the foreign stator genes and their effects on the motility of populations of cells in semi-solid agar. We considered causal mutations from two groups: they were either compatibilizing (compatibility group C), resulting in motility when non-motile initially, or motility-improving (compatibility group B), resulting in improved motility when motile immediately. We were interested in the diversity in causal mutations both within and between strains, their locations, the identity of mutations (SNPs, insertions or deletions) and their functional effects.

5.4.1 A comparison of causal mutations within strains

Our results showed the existence of incompatibilities that could be resolved by single mutations in the foreign stator genes. Furthermore, the discovered causal mutations provided insight into the predictability of evolution at the genetic level within strains, a longstanding problem in evolutionary theory [32,38–43]. All cBFM strains that harboured causal mutations in the stator genes, except for strain 11, showed at least one mutation that occurred in multiple replicates. In fact, ~65% of the mutations was not unique in this manner. This indicated that the first mutational steps in stator genes were constrained to follow a limited number of mutational routes. Although these results corroborated recent literature on the repeatability of evolution by natural selection [32,40–42], the fraction of identical mutations that were generated in several independent genotypes was higher in our experiment. The mentioned studies showed on average 6.4% of the total number of mutations to have been selected repeatedly, whereas we found an average fraction of 64.8%. This difference might be explained by the different nature of the experiments. While the experiments described in these studies [32,40–42] involved mutations that increased adaptation to a nutrient-limited environment, we focused on mutations that involved the incorporation of a new subunit into an already existing multi-protein complex, thereby potentially limiting the number of unique mutations. It is important to note that our experiment had a lower limit in terms of the smallest beneficial change that could be perceived as a flare, which may have led to an underestimation of the diversity of adaptive, first step changes available.

For strain 08 it might be possible that the first mutation comprised the A134S substitution in the MotA homologue in all strains, followed by either E87K (8.1d-P) or L284R (8.1b-P and 8.1c-P). Site-directed mutagenesis will reveal whether or not substitutions E87K and L284R have a selective advantage in isolation as well. If no selective advantage is found, it is likely that both mutations followed after A134S. However, if a selective advantage is found we will not be able to determine the original order of mutations.

5.4.2 A comparison of causal mutations between strains

In Chapters 2 and 4, we described that the mutations that resulted in improved motility were more diverse in terms of genes they affected than mutations resulting in compatibility. Here we found that all of the cBFM strains of compatibility group B, which harboured compatible stators, could evolve improved motility by either mutations in the foreign stator genes or at other loci. In contrast, the cBFM strains that contained incompatible stators (group C) evolved motility by either mutations in the foreign stator genes or at other loci—none of these strains used both mutational

routes. These findings are consistent with the idea that cBFM strains with incompatible stators were more constrained in terms of genes that afforded adaptive mutational routes to motility. In this case, the complexity of the conflicting interactions between stator and BFM/cell is higher for the stators that did not result in motility upon introduction.

Next, we compared the location of causal mutations between strains (Figure 5.1), as this might highlight strain-specific constraints. cBFM strains 11, 18 and 29 harboured causal mutations at a specific domain in only one of the stator genes. In contrast to these strains, we found mutations in multiple domains in strains 08 and 17. Strain 17 even gained mutations in both MotA and MotB homologues. Two mutations that might indicate a similar mechanism behind evolution of improved motility, were the plug-domain mutations in MotB homologues found in strains 11 and 17. These comparisons of mutations thus revealed that, in general, different domains were mutated in different strains. They furthermore showed that the foreign stators differed in the number of different domains in the MotA and MotB proteins that afforded mutational ways to resolve interaction issues with the host BFM and other cellular components. As the structure of the BFM can be vastly different for different bacterial species, both in terms of size and shape [44], we propose that the diversity in the causal stator mutations we observed was, at least in part, explained by structural/functional diversification of the BFMs in different species of bacteria. Although the underlying causes for specific mutations are not known, several possibilities for stator incompatibilities with the *E. coli* BFM exist. These include problems of docking the stator complex into the BFM, problems in aligning the MotA and FliG interaction sites and problems in binding to the peptidoglycan layer. Incompatibilities with the *E. coli* codon bias might additionally underlie the causal mutations in strain 29.

In Chapter 3 we proposed that big-benefit mutations were especially located in the stator genes and *fliG*, although not all mutations in these genes were expected to result in a big benefit. Furthermore, we proposed that strains 13, 18 and 29 lacked the availability of big-benefit mutations early in evolution (Chapter 2). Big-benefit mutations were defined as mutations that conferred an increase in relative population-swimming-speed of >50% of the difference between the highest speed possible in that strain and the initial speed. The results presented in this chapter showed that the identified causal mutations in the stator genes conferred a big-benefit only in strains 08, 11 and 17. This is consistent with the idea of the lack of big-benefit mutations early in evolution in the other three strains and it is also consistent with the idea that not all mutations in the stator genes confer a big benefit. The idea that not all mutations in the stator genes result in a big benefit, is further supported by the observation that the 18bp insertion in strain 17 and the A134S substitution in strain 08 confer only a small benefit, going from ATM to slow-motile and from non-motile to ATM, respectively.

5.4.3 Functional effects of causal mutations

Four different types of genetic mutations exist: Single Nucleotide Polymorphisms (SNPs), insertions, deletions and inversions. SNPs result in changes in single nucleotides only and often result in minor changes in protein structure and function. On the other hand, (large) insertions or deletions (InDels for short) or inversions can result in great alterations in protein structure and function and are therefore mostly extremely deleterious for the current function of a protein [30,31]. However, InDels can potentially play a major role in the evolution of new functions [31]. When examining the

discovered causal mutations, we only found InDels in strain 17 (in addition to SNPs in this strain). This indicated that SNPs in the stator genes were more common in the evolution of compatibility and improved motility. However, in contrast to the often-reported deleterious nature of InDels, the deletions in the MotB homologue of strain 17 resulted in a relative population-swimming-speed that was higher than most SNPs (Figure 5.2). This remarkable result indicated the potential value of InDels in improving the function of proteins [30,45].

Below, we examined the structural or functional implications of the causal mutations identified in this chapter. Starting from the SD-sequence of MotA, we found strain 29 to harbour a single mutation in this region. We expected this mutation to affect the protein levels of MotA. Although several SD sequences are used by *E. coli*, the most efficient sequence is AGGAGG [46]. In this instance, a possible SD sequence was changed from AGGATG to AGGATT, which is a poor SD-sequence in *Mycobacterium smegmatis* [47] and a reasonable SD-sequence in *Lactobacillus* [48]. Additional experiments are needed to reveal whether this mutation positively affected MotA protein levels.

The synonymous N2N substitution in strain 29, which is not favoured by the *E. coli* codon bias, might have resulted in increased protein levels. Although counter-intuitive, a recent study showed that rare codons at the N-terminus of proteins could lead to increased protein levels [49]. This, in turn, is a potential mechanism for improving the functioning of foreign proteins [13,50]. Given this result, we propose that the mutation in the SD-sequence of MotA also increased MotA protein levels.

The L3M mutation in strain 29, which represents the third mutation at the N-terminus of MotA and of which the putative function is not described in literature yet, might have caused a structural change, causing a better fit between the stator and the rotor. A similar role is proposed for the 18bp tandem repeat in the second periplasmic loop of MotA in strain 17, as it may have caused an extension of the third transmembrane helix or may have changed the length of the periplasmic loop. Alternatively, it might have affected the translocation of ions across the membrane, as the third and fourth transmembrane helices of MotA form the ion channel together with the transmembrane helix of MotB [17,51,52].

We propose that the four mutations in the cytoplasmic domains of MotA influenced the rotor-stator interaction. Substitution E87K in strain 08 might have resulted in a direct interaction with FliG, as it confers a charge reversal (negative to positive) in the vicinity of the MotA-FliG interaction site [17]. The other mutations are located in regions that were previously indicated to harbour mutations important for motility, possibly by affecting the alignment of the MotA-FliG interaction site [33,34]. As the L284R mutation is located close to the SD-sequence of MotB, it might have affected MotB-protein levels also.

The four mutations we discovered in MotB were predicted to either affect the flow of ions across the membrane or to result in an improved alignment between the stator and the rotor, as both functions are essential to the generation of torque by the BFM [17,53–56]. Before the stator is incorporated into the BFM, the plug domains of MotB homologues are involved in preventing the flow of ions across the membrane [26]. After incorporation, plug domains interact with each other [26] and thereby might influence the interaction between the PBD and the peptidoglycan. This, in turn, might influence the interaction between stator and rotor [26,27]. In light of motility of populations in semi-solid agar, mutations in the plug domain were expected to result in improved rotor-stator interaction, as leaky stators would likely cause cells to undergo growth arrest by

compromising the proton motive force [26]. Structural changes leading to improved rotor-stator interaction were also expected for the mutations in the C-terminal region of the transmembrane helix. However, changes in ion translocation were also possible, as the helix is part of the ion channel of the stator complex [17,51,52].

5.5 Conclusion

Taken together, the results showed that pre-adapted but non- or sub-functional foreign stators can be integrated via single-step mutations within the component itself. We showed that the locations in the stator proteins at which the causal mutations occurred, differed between strains, providing evidence in support of the idea that the BFM of the different donor species had diverged functionally in different directions. We found that for some stators, one-step compatibilization was possible via different mutations; however, the majority of causal mutations was observed in multiple replicate selection lines, indicating that the options for mutational integration within the component were limited. While the majority of causal mutations discovered in the stators involved single base pair changes, InDels were found occasionally. This is remarkable given the deleterious nature of most deletions [30,31]. Lastly, we showed that many, but not all, mutations in stator proteins were big-benefit mutations. This might indicate that the structure of epistatic constraint within our experimental system is such that the foreign component itself, rather than the accepting protein complex and cellular environment, contains the most options for large beneficial changes. Although detailed insight into the mechanisms by which the mutations caused integration remain to be elucidated, our results point at several structural/functional changes as well as regulatory effects.

Experimental data on the underlying mechanisms behind protein-complex evolution is scarce. In this chapter we provided some direct experimental insight into the creative potential of evolution to resolve inter-component conflicts at the compatibility horizon. Single mutations in pre-adapted foreign components that were severely dysfunctional could optimize their interaction with the host complex and cell systems. The foreign components in our experiment functioned in protein complexes with the same function as the host complex and had diverged to a point at which they became incompatible.

It is perhaps easy to see that this type of one-step integrating mutations, and the subsequent trajectories described in previous chapters, have played a role in adaptive evolution of the BFM by compositional evolution. One example is the case of a *Shewanella* species that is described in the introduction [57]. In contrast, at first sight, our observations appear to be a far cry from the events that facilitated *de novo* evolution of protein complexes. However, given the extraordinarily strong comparative evidence in support of a compositional origin of protein complexes [4–11], we must conclude that components sufficiently pre-adapted for co-option into functional protein complexes must have existed in the biosphere. If one accepts this, then it is likely that the trajectories for integration of the kind described in this thesis also facilitated *de novo* protein-complex evolution by pushing back the compatibility horizon and increasing the number of components available for compositional evolution.

5.6 Acknowledgements

We thank M. van 't Hof for performing the first experiments in the search for first-step mutations.

5.7 Materials and Methods

5.7.1 Ancestral strains and culturing conditions

The construction of the ancestral strains used in this chapter (Table 4.1) was described in detail in Chapter 2. The stator genes on the genome of *E. coli* K12 MG1655 were replaced by homologous recombination by a *rpsL-Neo* cassette [58], after which *rpsL* was replaced by *rpsL150*. Next, the foreign stator genes were cloned into and expressed from the low-copy plasmid pBAD33 [59], which also harboured a chloramphenicol-resistance gene. The strains that were isolated from primary flares (Chapter 2) and from which plasmid DNA was isolated to determine causality of mutations in stator genes are not listed in Table 5.6. This is because they were similar to the ancestral strains, except for mutations.

All strains used in this chapter were grown overnight before use (O/N; 5 µl from -80°C stock; 16-17h; 250rpm; 37°C; orbit diameter 2.5cm) in 5 ml liquid LB medium (20 g/l LB powder) supplemented with 0.0025% chloramphenicol (w/v), unless stated otherwise.

Table 5.6. Ancestral strains used in this chapter.

Strain ID	Host strain	Stator genes on plasmid pBAD33	Donor strain stator genes
Strain 01	<i>Escherichia coli</i> K12 str. MG1655 Δ <i>motAB</i> (short Δ <i>motAB</i>)	<i>motAB</i>	<i>Escherichia coli</i> K12 str. MG1655
Strain 08	Δ <i>motAB</i>	<i>lafTU</i>	<i>Escherichia coli</i> O111:H-str. 11128
Strain 11	Δ <i>motAB</i>	<i>motPS</i>	<i>Bacillus pseudofirmus</i> OF4
Strain 13	Δ <i>motAB</i>	<i>lafTU</i>	<i>Photobacterium profundum</i> SS9
Strain 17	Δ <i>motAB</i>	<i>motAB</i>	<i>Listeria monocytogenes</i> EGD-e
Strain 18	Δ <i>motAB</i>	<i>motPS</i>	<i>Bacillus megaterium</i> DSM319
Strain 29	Δ <i>motAB</i>	<i>motAB</i> "1"	<i>Rhodospirillum centenum</i> SW

5.7.2 Determination of causality of mutations

Causality of mutations was determined on semi-solid agar plates. Semi-solid agar plates consisted of 90mm petri dishes with 27 ml of 50% LB medium (v/v; Tryptone, 10 g/l; Yeast Extract, 2.5 g/l; NaCl, 10 g/l) supplemented with 0.3% agar (w/v), 0.2% L-arabinose (w/v) and 0.0025% chloramphenicol (w/v). Plates were prepared 19-20h in advance and were dried at room temperature in stacks of 3 plates. 5 µl from an O/N culture was inoculated in the centre of a semi-solid agar plate and plates were incubated at 37°C in stacks of three plates in airtight, plastic boxes (DBP, 500 ml). Motility of genotypes isolated from primary flares (*strains* in remainder of text) was scored qualitatively: the distance travelled after 24h by a population of cells was compared to the distance travelled by a population of cells of their direct, unevolved ancestor. When these strains travelled further than their

ancestor, we assumed these strains to be more motile and we isolated from these strains the plasmids on which the stator genes were located (Machery-Nagel Nucleospin® Plasmid QuickPure). Next, we transformed these evolved plasmids into *E. coli* cells lacking stator genes on the genome ($\Delta motAB$). When a population of these cells was more motile on semi-solid agar than a strain harbouring the unevolved, foreign stators, causal mutations had to be present on the plasmid. Causal mutations in the stator genes were identified by Sanger sequencing.

5.7.3 Relative population-swimming-speed

5 μ l from an O/N culture of a strain was inoculated in the centre of a semi-solid agar plate, which was prepared as described before. Semi-solid agar plates were incubated in stacks of 3 plates in airtight, plastic boxes (DPB; 500 ml) at 37°C. From 4h onwards, the distance the front of the population had travelled was measured every hour. In between two measurements, the plates were re-incubated at 37°C. As the travelled distance increased linearly with time [60], the population swimming speed could be calculated as $\frac{\Delta distance}{\Delta time}$. Unless stated otherwise, three measurement points were used for calculating the population swimming speed. Relative population-swimming-speed of a focal strain was calculated as $\frac{Speed\ focal\ strain}{Speed\ Strain\ 01}$.

5.7.4 Statistical analysis

In this chapter, parametric statistical analyses were performed. First, an overall one-way ANOVA was performed, followed by Student's *t*-tests. See text for details. When compensating for multiple comparisons, the Benjamini-Hochberg procedure was used [61].

5.8 Literature

- 1 Sowa Y & Berry RM (2008) Bacterial flagellar motor. *Q. Rev. Biophys.* **41**, 103–32.
- 2 Forgac M (2007) Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. *Nat. Rev. Mol. Cell Biol.* **8**, 917–29.
- 3 Finnigan GC, Hanson-Smith V, Stevens TH & Thornton JW (2012) Evolution of increased complexity in a molecular machine. *Nature* **481**, 360–4.
- 4 Pallen M & Matzke N (2006) From The Origin of Species to the origin of bacterial flagella. *Nat. Rev. Microbiol.* **4**, 784–790.
- 5 Liu R & Ochman H (2007) Stepwise formation of the bacterial flagellar system. *Proc. Natl. Acad. Sci.* **104**, 7116–7121.
- 6 Mulkidjanian A & Makarova K (2007) Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. *Nat. Rev. Microbiol.* **5**, 892–899.
- 7 Dolezal P, Likic V, Tachezy J & Lithgow T (2006) Evolution of the molecular machines for protein import into mitochondria. *Science* **313**, 314–8.
- 8 Clements A, Bursac D, Gatsos X, Perry AJ, Civciristov S, Celik N, Likic VA, Poggio S, Jacobs-Wagner C, Strugnell RA &

- Lithgow T (2009) The reducible complexity of a mitochondrial molecular machine. *Proc. Natl. Acad. Sci.* **106**, 15791–15795.
- 9 Archibald JM, Logsdon Jr. JM & Doolittle WF (2000) Origin and Evolution of Eukaryotic Chaperonins: Phylogenetic Evidence for Ancient Duplications in CCT Genes. *Mol. Biol. Evol.* **17**, 1456–1466.
- 10 Gabaldón T, Rainey D & Huynen M (2005) Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I). *J. Mol. Biol.* **348**, 857–70.
- 11 Seidl MF & Schultz J (2009) Evolutionary flexibility of protein complexes. *BMC Evol. Biol.* **9**, 155.
- 12 Watson RA (2002) *Compositional Evolution: Interdisciplinary Investigations in Evolvability, Modularity and Symbiosis*.
- 13 Lind PA, Tobin C, Berg OG, Kurland CG & Andersson DI (2010) Compensatory gene amplification restores fitness after inter-species gene replacements. *Mol. Microbiol.* **75**, 1078–89.
- 14 Omer S, Kovacs A, Mazor Y & Gophna U (2010) Integration of a foreign gene into a native complex does not impair fitness in an experimental model of lateral gene transfer. *Mol. Biol. Evol.* **27**, 2441–5.
- 15 Perica T, Chothia C & Teichmann S a (2012) Evolution of oligomeric state through geometric coupling of protein interfaces. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8127–32.
- 16 Hashimoto K & Panchenko AR (2010) Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20352–20357.
- 17 Blair DF (2003) Flagellar movement driven by proton translocation. *FEBS Lett.* **545**, 86–95.
- 18 Chun S & Parkinson J (1988) Bacterial motility: membrane topology of the Escherichia coli MotB protein. *Science* **239**, 276–278.
- 19 Khan S, Dapice M & Reese TS (1988) Effects of mot Gene Expression on the Structure of the Flagellar Motor at the Marine Biological Laboratory. *J. Mol. Biol.* **202**, 575–584.
- 20 de Mot R & Vanderleyden J (1994) The C-terminal sequence conservation between OmpA-related outer membrane proteins and MotB suggests a common function in both Gram-positive and Gram-negative bacteria, possibly in the interaction of these domains with peptidoglycan. *Mol. Microbiol.* **12**, 333–334.
- 21 Koebnik R (1995) Proposal for a peptidoglycan-associating alpha-helical motif in the C-terminal regions of some bacterial cell-surface proteins. *Mol. Microbiol.* **16**, 1269–1270.
- 22 Zhou J & Blair DF (1997) Residues of the Cytoplasmic Domain of MotA Essential for Torque Generation in the Bacterial Flagellar Motor. *J. Mol. Biol.* **273**, 428–439.
- 23 Lloyd SA & Blair DF (1997) Charged Residues of the Rotor Protein FlIG Essential for Torque Generation in the Flagellar Motor of Escherichia coli. *J. Mol. Biol.* **266**, 733–744.
- 24 Zhou J, Lloyd SA & Blair DF (1998) Electrostatic interactions between rotor and stator in the bacterial flagellar motor. *Proc. Natl. Acad. Sci.* **95**, 6436–6441.
- 25 Krogh A & Rapacki K (2013) <http://www.cbs.dtu.dk/services/TMHMM/> .
- 26 Hosking ER, Vogt C, Bakker EP & Manson MD (2006) The Escherichia coli MotAB proton channel unplugged. *J. Mol. Biol.* **364**, 921–37.
- 27 Garza A, Biran R, Wohlschlegel J & Manson M (1996) Mutations in motB suppressible by changes in stator or rotor components of the bacterial flagellar motor. *J. Mol. Biol.* **258**, 270–85.

- 28 Braun TF, Poulson S, Gully JB, Empey JC, Van Way S, Putnam a & Blair DF (1999) Function of proline residues of MotA in torque generation by the flagellar motor of *Escherichia coli*. *J. Bacteriol.* **181**, 3542–51.
- 29 Brown PN, Terrazas M, Paul K & Blair DF (2007) Mutational analysis of the flagellar protein FliG: sites of interaction with FliM and implications for organization of the switch complex. *J. Bacteriol.* **189**, 305–12.
- 30 Tóth-Petróczy A & Tawfik DS (2014) Hopeful (protein InDel) monsters? *Structure* **22**, 803–4.
- 31 Tóth-Petróczy A & Tawfik DS (2013) Protein insertions and deletions enabled by neutral roaming in sequence space. *Mol. Biol. Evol.* **30**, 761–71.
- 32 Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y & Palsson BØ (2009) Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome Biol.* **10**, R118.
- 33 Yakushi T, Yang J, Fukuoka H, Homma M & Blair DF (2006) Roles of Charged Residues of Rotor and Stator in Flagellar Rotation : Comparative Study using H⁺-Driven and Na⁺-Driven Motors in *Escherichia coli* Roles of Charged Residues of Rotor and Stator in Flagellar Rotation : Comparative Study using H⁺-Driven. *J. Bacteriol.* **188**, 1466–1472.
- 34 Hosking ER & Manson MD (2008) Clusters of charged residues at the C terminus of MotA and N terminus of MotB are important for function of the *Escherichia coli* flagellar motor. *J. Bacteriol.* **190**, 5517–21.
- 35 Casadesús J & Low D (2006) Epigenetic gene regulation in the bacterial world. *Microbiol. Mol. Biol. Rev.* **70**, 830–56.
- 36 Kumar R & Rao DN (2013) Role of DNA Methyltransferases in Epigenetic Regulation in Bacteria. *Subcell. Biochem.* **61**, 81–102.
- 37 Open WetWare *Escherichia coli* Codon Usage. .
- 38 Wichman HA (1999) Different Trajectories of Parallel Evolution During Viral Adaptation. *Science* **285**, 422–424.
- 39 Cooper V & Lenski R (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* **407**, 736–9.
- 40 Ostrowski EA, Woods RJ & Lenski RE (2008) The genetic basis of parallel and divergent phenotypic responses in evolving populations of *Escherichia coli*. *Proc. Biol. Sci.* **275**, 277–84.
- 41 Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR & Palsson BØ (2006) Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat. Genet.* **38**, 1406–12.
- 42 Woods R, Schneider D, Winkworth CL, Riley MA & Lenski RE (2006) Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9107–9112.
- 43 Brockhurst MA, Colegrave N & Rozen DE (2011) Next-generation sequencing as a tool to study microbial evolution. *Mol. Ecol.* **20**, 972–80.
- 44 Chen S, Beeby M, Murphy GE, Leadbetter JR, Hendrixson DR, Briegel A, Li Z, Shi J, Tocheva EI, Müller A, Dobro MJ & Jensen GJ (2011) Structural diversity of bacterial flagellar motors. *EMBO J.* **30**, 2972–81.
- 45 Arpino JAJ, Reddington SC, Halliwell LM, Rizkallah PJ & Jones DD (2014) Random Single Amino Acid Deletion Sampling Unveils Structural Tolerance and the Benefits of Helical Registry Shift on GFP Folding and Structure. *Structure* **22**, 889–898.
- 46 Vimberg V, Tats A, Remm M & Tenson T (2007) Translation initiation region sequence preferences in *Escherichia coli*.

BMC Mol. Biol. **8**, 100.

- 47 Unniraman S, Chatterji M & Nagaraja V (2002) A hairpin near the 5' end stabilises the DNA gyrase mRNA in *Mycobacterium smegmatis*. *Nucleic Acids Res.* **30**, 5376–81.
- 48 Vanderslice P, Copeland WC & Robertus JD (1986) Cloning and nucleotide sequence of wild type and a mutant histidine decarboxylase from *Lactobacillus* 30a. *J. Biol. Chem.* **261**, 15186–91.
- 49 Goodman DB, Church GM & Kosuri S (2013) Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science* **342**, 475–479.
- 50 Pettersson ME, Sun S, Andersson DI & Berg OG (2009) Evolution of new gene functions: simulation and analysis of the amplification model. *Genetica* **135**, 309–24.
- 51 Blair DF & Berg HC (1990) The MotA protein of *E. coli* is a proton-conducting component of the flagellar motor. *Cell* **60**, 439–49.
- 52 Stolz B & Berg HC (1991) Evidence for interactions between MotA and MotB, torque-generating elements of the flagellar motor of *Escherichia coli*. *J. Bacteriol.* **173**, 7033–7.
- 53 Manson MD, Tedesco PAT, Berg HC, Haroldt FM & van der Drift C (1977) A protonmotive force drives bacterial flagella. *Proc. Natl. Acad. Sci.* **74**, 3060–3064.
- 54 Zhu S, Takao M, Li N, Sakuma M, Nishino Y, Homma M, Kojima S & Imada K (2014) Conformational change in the periplamic region of the flagellar stator coupled with the assembly around the rotor. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13523–8.
- 55 Blair DF, Kim DY & Berg HC (1991) Mutant MotB Proteins in *Escherichia coli*. *J. Bacteriol.* **173**.
- 56 Togashi F, Yamaguchi S, Kihara MAY & Aizawa S (1997) An extreme clockwise switch bias mutation in flig of *Salmonella typhimurium* and its suppression by slow-motile mutations in motA and motB. *J. Bacteriol.* **179**, 2994–3003.
- 57 Paulick A, Koerdt A, Lassak J, Huntley S, Wilms I, Narberhaus F & Thormann KM (2009) Two different stator systems drive a single polar flagellum in *Shewanella oneidensis* MR-1. *Mol. Microbiol.* **71**, 836–50.
- 58 Heermann R, Zeppenfeld T & Jung K (2008) Simple generation of site-directed point mutations in the *Escherichia coli* chromosome using Red(R)/ET(R) Recombination. *Microb. Cell Fact.* **7**, 14.
- 59 Guzman LM, Belin D, Carson MJ, Beckwith J, Guzman L, Belin D & Carson MJ (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130.
- 60 Adler J (1966) Chemotaxis in bacteria. *Science* **153**, 708–716.
- 61 Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300.

5.9 Supplementary information

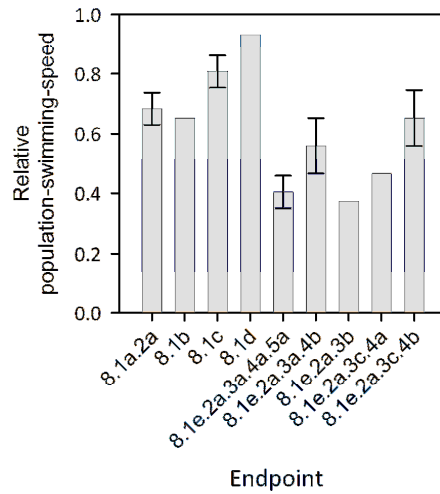


Figure 5.4. Population swimming speeds, relative to a strain harbouring wildtype *E. coli* stator genes, of the endpoints of strain 08. Endpoints are genotypes isolated from flares, that were not able to evolve improved motility during the selection experiment. $n = 3$. Error bars indicate standard deviation. The results for strains 8.1b and 8.1c suggest a small difference. Reproduced from Chapter 2.

6 Founder niche constrains evolutionary adaptive radiation

This chapter has been published as Régis C.E. Flohr, Carsten J. Blom, Paul B. Rainey and Hubertus J.E. Beaumont (2013) Founder niche constrains evolutionary adaptive radiation, *PNAS*, 110 (51), 20663-20668

6.1 Abstract

Adaptive radiation of a lineage into a range of organisms with different niches underpins the evolution of life's diversity. Although the role of the environment in shaping adaptive radiation is well established, theory predicts that the evolvability and niche of the founding ancestor are also of importance. Direct demonstration of a causal link requires resolving the independent effects of these additional factors. Here, we accomplish this using experimental bacterial populations and demonstrate how the dynamics of adaptive radiation are constrained by the niche of the founder. We manipulated the propensity of the founder to undergo adaptive radiation and resolved the underlying causal changes in both its evolvability and niche. Evolvability did not change, but the propensity for adaptive radiation was altered by changes in the position and breadth of the niche of the founder. These observations provide direct empirical evidence for a link between the niche of organisms and their propensity for adaptive radiation. This general mechanism may have rendered the evolutionary dynamics of extant adaptive radiations dependent on chance events that determined their founding ancestors.

6.2 Introduction

Rapid diversification of a single lineage into organisms with different niches - adaptive radiation - underpins the evolution of biodiversity [1–4]. Here, we define a niche as the complex of reciprocal ecological interactions between an organism and its environment that governs organismal fitness (following refs. [5–7]). Although the dynamics of adaptive radiation are determined by the interplay of many factors, two universal prerequisites are sufficiency of founder evolvability and availability of ecological opportunity [1–4,8–12]. Founder evolvability concerns the maximal range of derived organisms with different niches that can be accessed from the founding ancestor by mutation and recombination over an interval of evolutionary time [3,8,13–16]. Ecological opportunity for adaptive radiation emerges when the environment allows the possibility of invasion and persistence of multiple derived organisms with different niches [17]. Spectacular adaptive radiations, such as those of the Galapagos finches [18], African Rift Lake cichlids [15] or Caribbean Anolis lizards [19], can occur when evolvability and ecological opportunity interact to facilitate generation of niche specialists while precluding superior generalists, which might otherwise usurp ecological opportunity.

Since Simpson [2] formulated the requirements for adaptive radiation in the 1950s, the environment has been established as a major determinant of ecological opportunity [3]. This is supported by a large body of work on extant adaptive radiations [3,8,9,11,15,18,19]. In addition, evolutionary experiments with bacteria have provided direct evidence that links environmental components of ecological opportunity to patterns of adaptive radiation [20–30].

However, because ecological opportunity emerges from the interaction between the organism and the environment, it is also influenced by the niche of the (prospective) founder (e.g., positively by key innovations and negatively by resource usurpation [3]). Compared with the role of the environment, relatively little is known about the effects of the founder's niche on the dynamics of adaptive radiation. Comparative studies reveal differences in the tempo and outcome of parallel adaptive radiations that may have arisen from ecological differences among the founding ancestors [31,32]. In contrast, phylogenetic reconstruction of the niche of founders of extant adaptive radiations is yet to resolve significant correlations with patterns of diversification [4,8]. An effect of founder niche on evolutionary branching has been observed in experimental bacterial populations, demonstrating that founder specialization can increase the likelihood of diversification [33].

In addition to its niche, a second property of the founder that shapes adaptive radiation is evolvability [3,8,13–16]. Founder evolvability places an upper limit on the number of organisms with different niches that can emerge by adaptive radiation over an interval of evolutionary time. Moreover, for adaptive radiation to occur, founder evolvability must facilitate generation of diverse niche specialists but preclude “Darwinian demons” with superior, generalist niches that would impede adaptive radiation [10,34]. Studies on extant adaptive radiations have begun to probe the impact of evolvability [16,35,36], and its importance for diversification in general is supported by evidence from experimental evolution (e.g., refs. [37,38]).

The ecological theory of adaptive radiation predicts that founder evolvability and niche have an impact on the dynamics of diversification [8]. An important step toward further establishment of these factors is direct demonstration of their effects. To achieve this, it is necessary to manipulate the evolvability and niche of the founder experimentally and resolve their independent effects on adaptive radiation. Here, we accomplish this using experimental populations of the bacterium

Pseudomonas fluorescens SBW25. First, we exploit a previously described evolutionary selection regime to manipulate the propensity for adaptive radiation of the founder [39]. We then apply a combination of analyses to resolve the underlying causal changes in evolvability and niche. The results provide a direct demonstration of the founder-niche effect on the dynamics of adaptive radiation.

6.3 Results

When cultured in statically incubated microcosms, *P. fluorescens* populations rapidly diversify by mutation and selection into coexisting genotypes with different niches, which can be distinguished on the basis of the morphology of their colonies when grown on agar plates [22]. A critical difference between the niches that evolve is the mode of growth in static microcosms. The ancestral genotype, which forms colonies of the smooth (SM) morphology class, grows predominantly as planktonic cells in static microcosms. In contrast, genotypes that form colonies of the wrinkly spreader (WS) class form thick mats at the air-liquid interface of static microcosms [22,40]. The SM ancestor and a WS genotype have been shown to be capable of reciprocal invasion from rare in static microcosms [22]. However, in shaken microcosms that preclude mat formation, WS genotypes cannot invade populations of the SM ancestor from rare [40]. These observations show that SM and WS genotypes occupy different niches and coexist due to frequency-dependent fitness interactions and trade-offs [10,22,34]. Coexistence among WS genotypes has also been shown to be stabilized by negative frequency-dependent fitness interactions, indicating differences between their niches [41]. The genetic changes underlying this model of adaptive radiation have been established for various niche specialists at the nucleotide level [40,42–47].

6.3.1 Evolution of adaptive radiation impaired genotypes

We manipulated the propensity for adaptive radiation of founding genotypes using the selection regime of Buckling *et al.* [39]. This involved selection of a population founded with the ancestral genotype in static microcosms for six consecutive selection rounds (each lasting 7d), which were separated by population bottlenecks that allowed the passage of a single, numerically dominant individual with an SM colony morphology. The regime caused the evolution of genotypes with a SM colony morphology but a reduced propensity for adaptive radiation [39]. The mechanistic causes underlying this observation, were not determined (*Discussion*).

We founded six replicate populations with the ancestral genotype. Each round of selection involved ~10 generations, excluding population turnover due to cell death. As expected, the propensity for adaptive radiation during one round of selection decreased significantly with increasing rounds of selection (Figure 6.1A; Shannon's index of colony morphology diversity; Spearman's correlation test: $r_s = -0.443$, $n = 42$, $P < 0.001$). To identify genotypes with a significantly reduced propensity for adaptive radiation (AR^r), we compared the diversity produced during adaptive radiation of genotypes that had undergone six rounds of selection with that produced by the ancestral genotype (AR^a). With the exception of replicate line 1, all round 6 genotypes were significantly impaired in their propensity to undergo adaptive radiation (Figure 6.1B; Shannon's index

of colony morphology diversity; Bonferroni-corrected Student's *t*-tests or analysis of deviance: $n = 3$, $P < 0.031$ for all comparisons). AR^r genotypes produced between zero and five genotypes with new, heritably different colony morphologies compared with between four and seven genotypes produced by the AR^a genotype. In agreement with previous observations [39], all AR^r genotypes had an increased fitness relative to the AR^a genotype during one selection round of 7d in static microcosms (Figure 6.1C; Bonferroni-corrected Student's *t*-tests: $n = 3$, $P < 0.035$ for all comparisons). All AR^r genotypes formed colonies of the ancestral SM class when grown on agar plates (Figure 6.1D). The reduced propensity for adaptive radiation could be caused by (i) decreased evolvability, (ii) reduced ecological opportunity due to niche evolution, or (iii) a combination of changes in both factors.

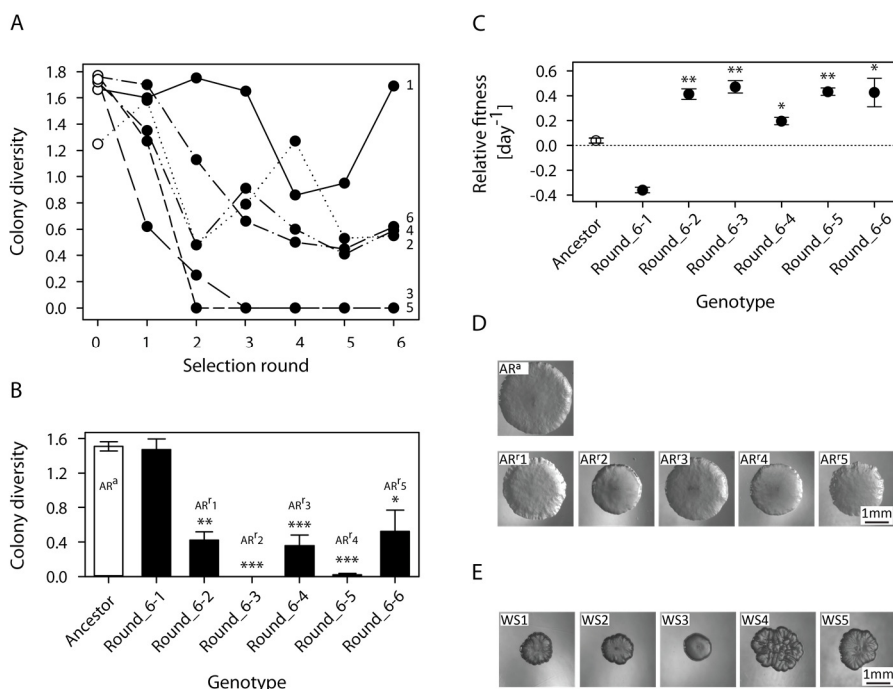


Figure 6.1. Evolution and characterization of AR^r genotypes. **A.** Colony morphology diversity (Shannon's diversity index) generated during adaptive radiation by genotypes that have undergone increasing rounds of selection for a reduced propensity to undergo adaptive radiation (numbers indicate replicate selection lines; $n = 1$). **B.** Replicate measurements of colony morphology diversity generated during adaptive radiation by the ancestor (AR^a) and six genotypes from round 6 ($n = 3$). **C.** Fitness of ancestral and round 6 genotypes relative to the ancestral genotype ($n = 3$). Bars indicate SEM. Asterisks indicate significant deviation from the ancestor (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; statistics are provided in main text). **D.** Colonies of the ancestral AR^a and AR^r genotypes. **E.** Colonies of five randomly selected, different AR^a-derived WS genotypes.

6.3.2 Quantifying genetic constraints

We first evaluated the hypothesis that adaptive radiation was suppressed by a change in evolvability that limited the mutational accessibility of genotypes with new niches. This required quantification of the rates at which these genotypes arise by random mutation in growing

populations of the AR^a and AR^r genotypes. However, the frequency of mutants in a population is also affected by natural selection, including selection engendered by the niche of the ancestral genotype. Therefore, to disentangle these factors, we measured the rate at which WS genotypes are generated by random mutation in the absence of the confounding influence of natural selection. This was achieved using a genetically encoded morph selection cassette that allows quantification of WS genotypes at low frequency [29,44] (*Materials and Methods*). WS genotypes represent $44.6 \pm 4.6\%$ of the diversity that evolves during adaptive radiation of the AR^a genotype ($n = 18$, mean \pm SEM; 7d radiation). Using a fluctuation test design, we estimated the probability with which the AR^a genotype and five AR^r genotypes mutated to WS genotypes upon cell division on the basis of the Ma-Sandri-Sarkar maximum likelihood estimator method [48,49]. Although two of five AR^r genotypes did not produce detectable WS genotypes during adaptive radiation (Figure 6.1B; AR^r2 and AR^r4), all AR^r genotypes had nonzero mutation rates to WS genotypes (Figure 6.2). Pairwise comparison of the mutational accessibility of WS genotypes from AR^a and AR^r genotypes did not support the hypothesis that adaptive radiation of AR^r genotypes was suppressed due to a reduction in the mutational accessibility of alternative niche-specialist genotypes. Rather, the accessibility of WS genotypes was significantly higher from all AR^r genotypes than from the AR^a genotype (Bonferroni-corrected Student's *t*-tests on estimates of mean number of mutations (m) in 40 populations: $P < 0.001$ for all comparisons; *Materials and Methods*).

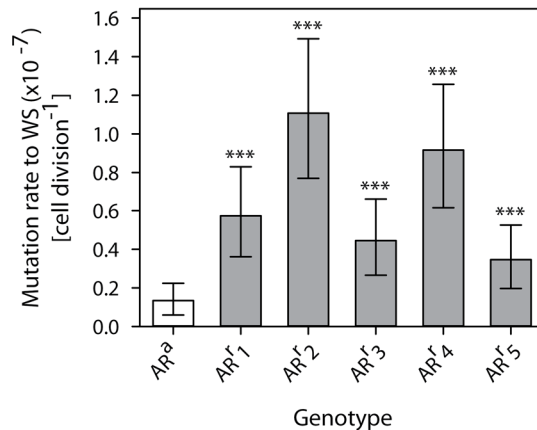


Figure 6.2. Mutation rates of ancestral (AR^a) and AR^r genotypes to WS genotypes per cell division (fluctuation tests; $n = 40$). Bars indicate 95% confidence intervals. Asterisks indicate significant deviation from the AR^a genotype (** $P < 0.001$; statistics are provided in main text).

6.3.3 Screening for altered ecological constraints

Next, we tested the hypothesis that the reduced propensity for adaptive radiation of AR^r genotypes was due to changes in their niche that limited the ecological opportunity for invasion of derived niche-specialist genotypes. To examine this, we measured the capacity of AR^r-derived WS genotypes, which were isolated using the morph selection cassette, to invade AR^a and AR^r genotypes

from rare (starting frequency of $\sim 10^{-5}$). Overall, the invasion fitness of AR^r-derived WS genotypes was positive but significantly lower against AR^r genotypes than against the AR^a genotype (Figure 6.3; paired Student's *t*-test: $n = 5$, $t = -3.74$, $P = 0.02$). These findings demonstrate that AR^r genotypes suppress invasion of derived WS genotypes through ecological interactions. Consequently, the observed reduction in WS genotype diversity during AR^r adaptive radiations must have involved changes in the niche of AR^r.

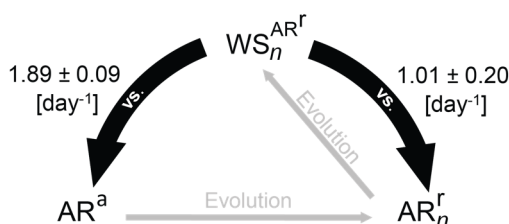


Figure 6.3. Invasion fitness of AR^r-derived WS genotypes (WS^{AR^r}) in populations of the AR^a genotype and their cognate AR^r ancestors. The term n indicates five different AR^r genotypes and cognate-derived WS genotypes. Values are means of the invasion fitness measurements of the five AR^r genotypes \pm SEM. The difference between mean invasion fitness values is significant ($n = 5$; statistics are provided in main text).

6.3.4 Quantifying niche evolution

How did the niche of AR^r genotypes change to constrain their potential for adaptive radiation? Despite the (deceiving) simplicity of *P. fluorescens* experimental radiations, not all ecological interactions that govern the dynamics have yet been elucidated. To quantify the niche of a genotype, we measured how it affected the invasion fitness of two of the niche-specialist genotypes that had been characterized thus far: the ancestral AR^a genotype that predominantly shows planktonic growth and a randomly selected AR^a-derived reference WS genotype (WS_{ref}) that grows predominantly in mats [22,40]. Thus quantified, the niche of any genotype can be represented as a point in a 2D niche space. The invasion assays were performed for 2d rather than 7d to avoid confounding effects associated with the evolution of new genotypes with alternative niches that occurs during 7d (approximately 10 generations, excluding cell turnover).

We used this approach to examine three hypotheses regarding the relative positions in niche space of the AR^a genotype, the five AR^r genotypes, and five randomly selected, different AR^a-derived WS genotypes (designated WS_1 – WS_5 ; Figure 6.1D and E). First, based on the SM/WS relationship described previously, we expected that the niches of the AR^a genotype and the WS_1 – WS_5 genotypes would occupy different positions in niche space. Second, our finding that AR^r genotypes had an increased potential to suppress mutationally-derived WS genotypes led us to test the hypothesis that the niches of AR^r genotypes had shifted toward those of WS_1 – WS_5 relative to the niche of the AR^a genotype. Finally, if we assume that the SM/WS trade off described previously is evolutionarily unbreakable, we expect that evolution of an increased ability to suppress invading WS_{ref} would be paralleled by a decrease of equal or larger magnitude in the ability to suppress AR^a, and *vice versa* (i.e. linear or strong trade-offs, respectively, that maintain or increase niche specialization). We

evaluated this by testing for deviation from a linear trade off drawn through the location of the AR^a genotype in niche space (Figure 6.4A, dashed line).

As expected, WS1–WS5 suppressed the WS_{ref} genotype more than it was suppressed by the AR^a genotype (Figure 6.4A; analysis of covariance followed by Bonferroni-corrected planned comparisons with AR^a: covariate = *invader_density*, $n = 5$, $F_{5,23} = 10.00$, $P < 0.003$ for all comparisons). The degree of suppression of the AR^a genotype varied substantially between the different genotypes WS1–WS5. Although some suppressed AR^a to a level consistent with a linear SM/WS trade off, WS2 deviated from this significantly (Figure 6.4A, linear trade off is indicated by the dashed line; Bonferroni-corrected Student's *t*-tests: $n = 5$, $t(8) = 5.11$, $P = 0.009$; *Materials and Methods*). All AR^r genotypes had an increased capacity to suppress invasion by the WS_{ref} genotype relative to the AR^a genotype (Figure 6.4A; analysis of covariance with Bonferroni-corrected planned comparisons with AR^a: covariate = *invader_density*, $n = 5$, $F_{5,23} = 8.16$, $P < 0.002$ for all comparisons). This demonstrates that the niches of all AR^r genotypes had significantly shifted toward that of WS genotypes. One of the five AR^r genotypes deviated significantly from a linear SM/WS tradeoff [AR^r3; Figure 6.4A; Bonferroni-corrected Student's *t*-tests: $n = 5$, $t(8) = 7$, $P = 0.002$]. This analysis indicates that the niche of AR^r3 (and also that of WS2) had become broader (i.e. generalized relative to the niche of the AR^a genotype). The finding that the WS_{ref} genotype was suppressed more by the AR^r genotypes than by the AR^a genotype shows that the effect of AR^r genotypes on invasion of AR^r-derived WS genotypes shown in Figure 6.3 also holds for AR^a-derived WS genotypes.

The salient characteristic of the niche of WS genotypes is formation of a thick mat at the air-liquid interface [22,40]. To examine if the increased overlap of the niches of AR^r genotypes with the WS niche involved increased exploitation of the air-liquid interface, we measured the amount of biomass in mats produced by the AR^a genotype, AR^r1–AR^r5 and WS1–WS5. This was achieved by collection of the mat from microcosms after 48h, suspension by vigorous mixing, and measurement of the turbidity (*Materials and Methods*). WS genotypes produced significantly thicker mats than the AR^a genotype, which produced very thin, difficult-to-discriminate mats (Figure 6.4B; Kruskal-Wallis test followed by Bonferroni-corrected planned comparisons with AR^a: $n = 5$, $W_s = 15$, $z = -2.514$, $P = 0.04$ for all comparisons). Mats of the AR^r genotypes appeared thinner than those of WS genotypes; nonetheless, all AR^r mats contained significantly more biomass than those formed by the AR^a genotype (Figure 6.4B; Kruskal-Wallis test followed by Bonferroni-corrected planned comparisons with AR^a: $n = 5$, $W_s = 15$, $z = -2.507$, $P = 0.04$ for all comparisons). All AR^r mats only contained genotypes with a SM colony morphology (~250 cells checked per genotype).

WS mats are reinforced by increased expression of extracellular cellulose [40]. Qualitative estimation of cellulose expression on agar plates containing Congo red, a cellulose-binding dye, did not indicate that AR^r cells had increased cellulose expression levels compared with AR^a cells.

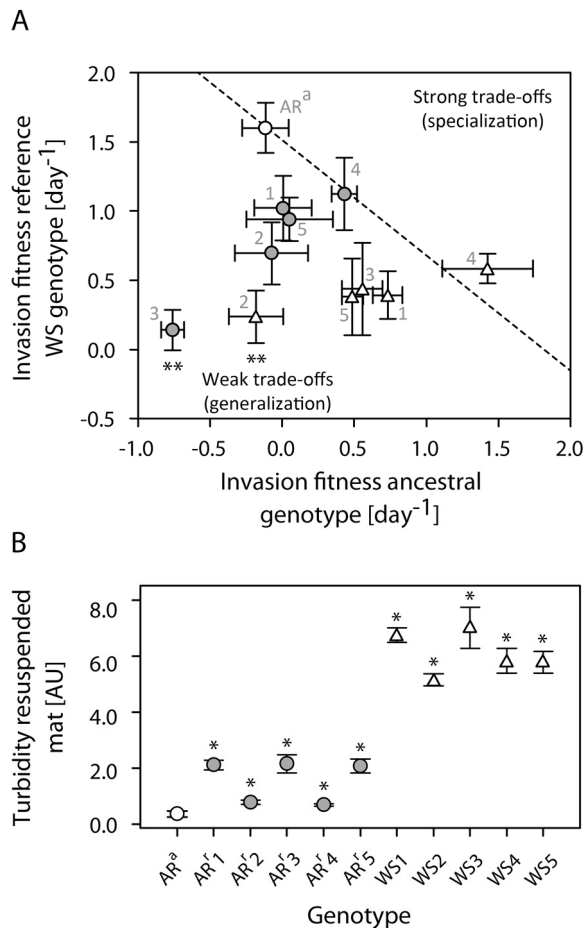


Figure 6.4. Niche characterization of evolved genotypes. **A.** Niches of AR^a (white circle), WS genotypes (white triangles), and AR^r genotypes (gray circles) quantified by invasion fitness assays with the AR^a and a reference WS genotype. Bars indicate SEM ($n = 5$). Grey numbers indicate replicate selection lines. The dashed line indicates niche evolution under a linear trade off from the ancestral niche. Asterisks indicate significant deviation from the linear trade off (** $P < 0.01$; statistics are provided in main text). **B.** Size of mats produced by AR^a (white circle) and AR^r (gray circles) genotypes and by five WS genotypes (white triangles) determined by measurement of turbidity after mat suspension. Bars indicate SEM ($n = 5$). Asterisks indicate significant deviation from AR^a (* $P < 0.05$; statistics are provided in main text). AU, arbitrary units.

6.3.5 Extended adaptive radiation

To examine how the altered properties of AR^r genotypes affected adaptive radiation in the longer term, we compared the evolutionary dynamics of AR^a and AR^r genotypes in populations that were propagated by serial dilution [22]. To allow us to interpret the dynamics in the same framework as the niche characterization, we used a selection regime that also involved selection rounds of 2d. Comparison of the mean diversity in populations founded with AR^a and AR^r genotypes over time revealed that the tempo of AR^r radiations was significantly lower (Figure 6.5; Shannon's index of

colony morphology diversity; Wilcoxon's rank-sum tests per time point: $n = 5$, $W_s \leq 17.0$, $z \leq -2.193$, $P < 0.05$ at selection rounds 0, 1, 2, and 3). AR^a and AR^r genotypes reached a plateau at a similar diversity by the fifth selection round.

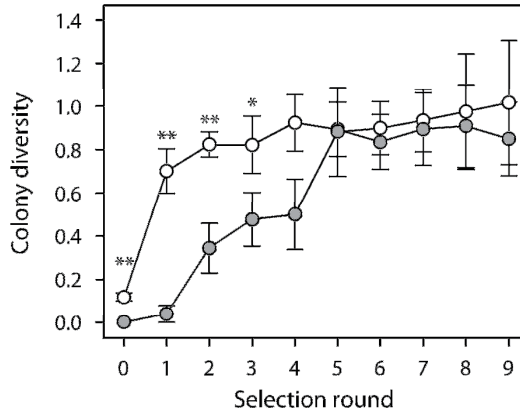


Figure 6.5. Mean colony morphology diversity (Shannon's diversity index) during longer term adaptive radiation of five replicate populations founded with AR^a genotypes (white circles) and five different AR^r genotypes (gray circles). Bars indicate SEM. Asterisks indicate significant deviation from AR^a (* $P < 0.05$, ** $P < 0.01$; statistics are provided in main text).

6.4 Discussion

We used experimental evolution of the bacterium *P. fluorescens* as a model to examine the link between the evolvability and niche of organisms and their propensity for evolutionary diversification by adaptive radiation. We accomplished this by decomposing organismal evolutionary constraints into two causative factors, namely, evolvability and niche. Our results provide experimental support for the theoretical prediction that the niche of the founder affects the dynamics of adaptive radiation [8].

The conclusion that the tempo of adaptive radiation was reduced as a result of changes in the niche of the founder is based on the observation that invading mutationally-derived WS genotypes were suppressed more by AR^r founders. Three additional observations support this conclusion. First, quantification of the evolvability of AR^r genotypes did not reveal altered genetic constraints that limit diversification; rather, mutational accessibility of the WS phenotype had increased. Second, the niche of AR^r genotypes had shifted toward that of WS genotypes and, in one case, had broadened relative to the ancestral genotype, which has a higher propensity for adaptive radiation (AR^a). The niches of the reference genotypes used to characterize niche evolution most likely do not represent all dimensions of the niche space of this model of adaptive radiation. Nonetheless, our analysis revealed differences that, at least in part, explain how AR^r genotypes limited the ecological opportunity available for their own adaptive radiation. Third, the finding that AR^r genotypes exploited the air-liquid interface more than the AR^a genotype shows that their niche shifted toward that of WS genotypes.

In addition to these general findings, our results shed new light on the *P. fluorescens* model of adaptive radiation. Discovery of *P. fluorescens* adaptive radiation and its extensive use in experimental evolutionary research have been facilitated by the fortuitous link between niche and colony morphology. Although our data support this pleiotropic link, they also indicate that shifts in the niche of genotypes with a SM colony morphology toward the WS niche are not necessarily paralleled by a shift from the SM morphology class to the WS morphology class. SM and WS genotypes appear to correspond to two potentially overlapping clusters in niche space. Recent work indicates the same holds for the niches of WS and fuzzy spreader genotypes [47].

Our quantification of the niches of AR^a, AR^r, and WS genotypes revealed evidence suggesting that increased competitiveness in the WS niche is not necessarily traded off in a one-to-one manner for a decrease in the ability to compete in the niche of the ancestral SM genotype. Within the niche space sampled by our approach, AR^r3 and WS2 deviated from linear and strong SM/WS trade-offs, evolving a broader niche than their ancestor. Our observation that niche generalization coincided with a reduced propensity for adaptive radiation is consistent with previous experimental work showing that specialization increases the likelihood of evolutionary branching in experimental bacterial populations [33].

We manipulated the propensity for adaptive radiation of the founder using an evolutionary selection regime described previously [39]. This previous study did not dissect the underlying causes. Rather, making a number of reasonable but untested assumptions, the authors concluded that reduced capacity for diversification was a consequence of altered genetic constraints engendered by increased adaptation of the founder to the ancestral SM niche. Our observations show that these assumptions and the conclusion they support do not hold.

More than a century ago, H. F. Osborn coined the term “adaptive radiation” to refer to the “differentiation of habit in several directions from a primitive type” that occurs during evolution [1]. Since then, our understanding of adaptive radiation (in the modern definition) has greatly increased, but some dynamics remain hidden in evolutionary history. At the cost of disregarding relevant complexity, our minimal experiment has empirically established a link between the niche of organisms and their propensity for adaptive radiation.

Rapid diversification is one of the hallmarks of adaptive radiation. AR^r genotypes diversified at a rate that was approximately threefold lower than that of the AR^a genotype. If the niches of the potential founders of natural adaptive radiations were diverse, and selection of the actual founding ancestors sufficiently stochastic, the relation between niche and diversification may have contributed to random variation in the tempo of extant adaptive radiations. The importance of this general principle as a cause of historical contingency in adaptive radiations is worthy of further study [50,51].

6.5 Acknowledgements

We thank A. Gunn and F. Jesse for preliminary experiments; J. van den Heuvel, D. de Ridder, and B. J. Zwaan for advice on statistical analyses; I. Geuzebroek, M. Gray, E. van Rijn, and I. Westerlaken for technical assistance; A. Buckling, T. F. Cooper, T. Fukami, and C. Kost for discussion; and R. Hermsen, F. Hol, S. Rossell, and three anonymous reviewers for constructive comments on the

manuscript. H.J.E.B. was supported by the Marsden Fund Council from government funding administered by the Royal Society of New Zealand and a Veni Fellowship from The Netherlands Organization for Scientific Research.

6.6 Materials and Methods

6.6.1 Selection of Adaptive Radiation Impaired Genotypes

P. fluorescens SBW25 (22) was grown in 30 mL glass vials filled with 6 mL of King's medium B (glycerol, 10 mL·L⁻¹; K₂HPO₄, 1.5 g·L⁻¹; MgSO₄·7H₂O, 1.5 g·L⁻¹; proteose peptone, 20 g·L⁻¹) supplemented with 0.0024% (w/v) vitamin B5 (D-pantothenic acid hemicalcium salt). Replicate microcosms were inoculated with ~10⁷ cells from -80°C glycerol stocks and incubated statically with loose caps at 28°C. After 7d, microcosms were mixed by vortexing, diluted, and plated on King's medium B agar plates (glycerol, 10 mL·L⁻¹; K₂HPO₄, 1.5 g·L⁻¹; MgSO₄·7H₂O, 1.5 g·L⁻¹; tryptone, 20 g·L⁻¹; agar, 15 g·L⁻¹) that were incubated for 2d at 28°C. Colony diversity was scored by counting the number of distinct, heritable colony morphology types, sub-SM and sub-WS types and the fuzzy spreader type [22], in a sample of 100 randomly selected colonies. Colony diversity was expressed using Shannon's index, $H' = -\sum_{i=1}^R p_i \ln(p_i)$, where R is the total number of different colony morphologies in a population and p_i is the frequency of the i-th colony morphology in the population. Cells from one numerically dominant colony of the SM class were grown for 16h in shaken microcosms with loose caps (28°C; 200rpm; orbit diameter 2.5cm; 30°angle) and stored at -80°C. New rounds were started with ~10⁷ SM cells from frozen stocks. This procedure was repeated five times.

6.6.2 Propensity for Adaptive Radiation

Genotypes were allowed to diversify for 7d in replicate microcosms, and colony diversity was analyzed as described above.

6.6.3 Fitness Assays

Cells of the genotype to be analyzed and labelled cells of the ancestral genotype (genomic *lacZ* [52]) were preconditioned by growth in shaken microcosms as described above. Competitions were founded with ~5 × 10⁶ cells of each competitor in a fresh microcosm and incubated statically for 7d as described above. Start and end population sizes of each competitor were determined by plating of a diluted sample on King's medium B plates containing X-Gal (0.4 g·L⁻¹) and scoring of at least 50 colonies. Fitness was expressed as the selection rate constant, $r_{ij} = \frac{\ln\left(\frac{N_i(t)}{N_i(0)}\right) - \ln\left(\frac{N_j(t)}{N_j(0)}\right)}{t}$, where N is the population size of genotypes to be analyzed (i) and the ancestral genotype (j) at the start (t = 0) and at the end (t = t) [53]. Genotypes with a fitness equal to that of the ancestor have a selection rate coefficient of zero.

6.6.4 Fluctuation Tests

The morph selection cassette comprises a kanamycin resistance gene downstream of a promoter (P_{wss_operon}) that is up-regulated in WS cells, rendering these cells kanamycin resistant [29,44]. Morph selection cassettes were introduced into the genotypes to be analyzed by triple mating and single cross-over into the chromosome. Fluctuation tests were performed using 40 replicate populations per genotype that were founded with ~ 20 cells in 50 μL of King's medium B with vitamin B5. Populations were grown in 96-well plates at 28°C (800 rpm, 2-mm orbit diameter) to a size of $1.35 \times 10^7 \pm 1.33 \times 10^6$ (mean population size of all populations of all genotypes \pm SD) using the turbidity measured at 500 nm and calibration curves. The number of WS mutants was determined by plating each population in its entirety on King's medium B plates supplemented with kanamycin (30 $\mu\text{g}\cdot\text{mL}^{-1}$) and Congo red (cellulose dye for detection of WS-specific cellulose overexpression; 15 $\text{mg}\cdot\text{L}^{-1}$) and subsequent counting of the number of red colonies with a WS morphology. The average number of mutation events per culture, m , was estimated using the Ma-Sandri-Sarkar maximum likelihood estimator method of the Fluctuation Analysis Calculator web tool [49]. Statistical comparisons of m values of different genotypes using Student's t -tests were performed according to the method of Rosche and Foster [54]. Per division mutation rates were calculated by dividing m by the total number of cell divisions per culture [$N(t) - 20$]. Variation in $N(t)$ between populations did not affect the statistical significance of mutation rate comparisons, as determined by estimating the effect of variation in $N(t)$ on m using the "p0" method [55].

6.6.5 AR^r-derived WS Invasion Fitness

Low frequencies of WS invaders were measured using the morph selection cassette-encoded selection marker (kanamycin). Resident and WS-invader cells were preconditioned as for fitness assays. Invasions were founded with a mixture of ~ 10 invader cells and $\sim 10^6$ resident cells and were incubated and analyzed in the same manner as the fitness assays. Start and end population sizes of resident and invader cells were measured by plating on regular King's medium B plates and the same plates with added kanamycin and Congo red (as for the fluctuation test). Invasion fitness was expressed as the selection rate coefficient calculated in the same manner as for the fitness assays.

6.6.6 Niche Quantification

Cells of the genotype to be analyzed (resident), labelled cells of the ancestor (invader; genomic *lacZ* [52]), and a randomly selected "reference" WS genotype (invader; derived from the *lacZ*-labeled ancestor after a 7d adaptive radiation) were preconditioned as for the fitness assays. Invasions were founded by a mix of $\sim 10^4$ invader cells and $\sim 10^6$ resident cells and incubated statically for 2d as in the selection experiment. Invasion fitness was determined by measuring the start and end population sizes of the resident and invader cells and was calculated in the same manner as for the fitness assays. Deviation of evolved genotypes from a linear SM/WS trade off was assessed by calculating the *invasion_fitness_ancestral_genotype*-value that would be expected under a linear trade off from the ancestral genotype on the basis of the measured *invasion_fitness_reference_WS_genotype*-value of the evolved genotypes, and subsequently testing the hypothesis that their expected and measured

invasion_fitness_ancestral_genotype-values are equal, in which case there was no deviation from the linear trade off.

6.6.7 Mat Size

Microcosms were founded as described for the selection of adaptive radiation impaired genotypes but incubated for 2d. Mats were collected from the air-liquid interface using a cut-off 1mL pipette tip. The collected material contained the mat material (fragmented) but also planktonic cells. Mat fragments were separated from the planktonic cells by washing the collected material in 30mL of Ringer's solution (NaCl, 8.6 g·L⁻¹; KCl, 0.3 g·L⁻¹; CaCl₂, 0.33 g·L⁻¹). From this mixture, a very small mat fragment was taken for analysis of the colony morphology of the cells it comprised by plating on King's medium B agar plates. Next, all mat fragments were collected from this mixture with a 1mL pipette tip and washed once by centrifugation and resuspension in Ringer's solution. The same was done for a 1mL sample of the remaining planktonic cells from this mixture. The amount of biomass in such collected mat- and planktonic-cell fractions was quantified by measuring the turbidity at 600nm after vortexing. Mat size was expressed as turbidity corrected for the contribution of the planktonic cells.

6.6.8 Extended Adaptive Radiation

Populations of AR^a and AR^r genotypes were founded in microcosms with ~10⁶ cells and incubated as described for the selection experiment. After 2d, microcosms were mixed by vortexing and frozen in -150°C glycerol stocks. Populations were propagated by transfer of 0.01% of the volume of the original population to fresh microcosms, after which the frozen population was stored at -80°C. This procedure was repeated nine times.

6.7 Literature

- 1 Osborn HF (1902) The Law of Adaptive Radiation. *Am. Nat.* **36**, 353–363.
- 2 Simpson G (1953) *The Major Features of Evolution* (Columbia Univ Press, New York).
- 3 Losos JB (2010) Adaptive radiation, ecological opportunity, and evolutionary determinism. *Am. Nat.* **175**, 623–39.
- 4 Glor RE (2010) Phylogenetic Insights on Adaptive Radiation. *Annu. Rev. Ecol. Evol. Syst.* **41**, 251–270.
- 5 McInerney GJ & Etienne RS (2012) Pitch the niche - taking responsibility for the concepts we use in ecology and species distribution modelling. *J. Biogeogr.* **39**, 2112–2118.
- 6 McInerney GJ & Etienne RS (2012) Stitch the niche - a practical philosophy and visual schematic for the niche concept. *J. Biogeogr.* **39**, 2103–2111.

- 7 McNerny GJ & Etienne RS (2012) Ditch the niche - is the niche a useful concept in ecology or species distribution modelling? *J. Biogeogr.* **39**, 2096–2102.
- 8 Schluter D (2000) *The Ecology of Adaptive Radiation* (Oxford Univ Press, Oxford).
- 9 Wagner CE, Harmon LJ & Seehausen O (2012) Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* **487**, 366–9.
- 10 Kassen R (2009) Toward a general theory of adaptive radiation: insights from microbial experimental evolution. *Ann. N. Y. Acad. Sci.* **1168**, 3–22.
- 11 Gillespie RG & Clague D (2009) *Encyclopedia of Islands* (Univ of California Press, Berkeley, CA).
- 12 Gavrillets S & Vose A (2005) Dynamic patterns of adaptive radiation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 18040–5.
- 13 Kirschner M & Gerhart J (1998) Evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8420–8427.
- 14 Conrad M (1990) The geometry of evolution. *Biosystems* **24**, 61–81.
- 15 Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. *Proc. Biol. Sci.* **273**, 1987–98.
- 16 Gavrillets S & Losos JB (2009) Adaptive Radiation: Contrasting theory with data. *Science* **323**, 732–737.
- 17 Yoder JB, Clancey E, Des Roches S, Eastman JM, Gentry L, Godsoe W, Hagey TJ, Jochimsen D, Oswald BP, Robertson J, Sarver B a J, Schenk JJ, Spear SF & Harmon LJ (2010) Ecological opportunity and the origin of adaptive radiations. *J. Evol. Biol.* **23**, 1581–96.
- 18 Grant PR & Grant BR (2011) *How and Why Species Multiply* (Princeton Univ Press, Princeton).
- 19 Losos JB (2009) *Lizards in an Evolutionary Tree* (Univ of California Press, Berkeley, CA).
- 20 Helling R, Vargas CN & Adams J (1987) Evolution of *Escherichia coli* during growth in a constant environment. *Genetics* **358**, 349–358.
- 21 Rozen DE & Lenski RE (2000) Long-term experimental evolution in *Escherichia coli* VIII. Dynamics of a Balanced Polymorphism. *Am. Nat.* **155**, 24–35.
- 22 Rainey PB & Travisano M (1998) Adaptive radiation in a heterogeneous environment. *Nature* **32**, 69–72.
- 23 Buckling a, Kassen R, Bell G & Rainey PB (2000) Disturbance and diversity in experimental microcosms. *Nature* **408**, 961–4.
- 24 Kassen R, Llewellyn M & Rainey PB (2004) Ecological constraints on diversification in a model adaptive radiation. *Nature* **431**, 984–989.
- 25 Maharjan R, Seeto S, Notley-McRobb L & Ferenci T (2006) Clonal adaptive radiation in a constant environment. *Science* **313**, 514–517.
- 26 Habets MGJL, Rozen DE, Hoekstra RF & de Visser JAGM (2006) The effect of population structure on the adaptive radiation of microbial populations evolving in spatially structured environments. *Ecol. Lett.* **9**, 1041–8.

- 27 Craig MacLean R, Dickson A & Bell G (2004) Resource competition and adaptive radiation in a microbial microcosm. *Ecol. Lett.* **8**, 38–46.
- 28 Brockhurst M a, Colegrave N, Hodgson DJ & Buckling A (2007) Niche occupation limits adaptive radiation in experimental microcosms. *PLoS One* **2**, e193.
- 29 Fukami T, Beaumont HJE, Zhang X-X & Rainey PB (2007) Immigration history controls diversification in experimental adaptive radiation. *Nature* **446**, 436–9.
- 30 Bailey SF, Dettman JR, Rainey PB & Kassen R (2013) Competition both drives and impedes diversification in a model adaptive radiation. *Proc. Biol. Sci.* **280**, 20131253.
- 31 Price T, Lovette IJ, Bermingham E, Gibbs H & Richman A (2000) The imprint of history on communities of North American and Asian warblers. *Am. Nat.* **156**, 354–367.
- 32 Ord T (2012) Historical contingency and behavioural divergence in territorial *Anolis* lizards. *J. Evol. Biol.* **25**, 2047–2055.
- 33 Spencer CC, Tyerman J, Bertrand M & Doebeli M (2008) Adaptation increases the likelihood of diversification in an experimental bacterial lineage. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 1585–9.
- 34 Remold S (2012) Understanding specialism when the Jack of all trades can be the master of all. *Proc. Biol. Sci.* **279**, 4861–9.
- 35 Lanfear R, Ho S, Love D & Bromham L (2010) Mutation rate is linked to diversification in birds. *Proc. Natl. Acad. Sci.* **107**, 20423–20428.
- 36 Kellermann V, van Heerwaarden B, Sgrò C & Hoffmann A (2009) Fundamental evolutionary limits in ecological traits drive *Drosophila* species distributions. *Science* **325**, 1244–1246.
- 37 Blount ZD, Barrick JE, Davidson CJ & Lenski RE (2012) Genomic analysis of a key innovation in an experimental *Escherichia coli* population (supplement). *Nature* **489**, 513–8.
- 38 Jasmin J-N & Kassen R (2007) Evolution of a single niche specialist in variable environments. *Proc. Biol. Sci.* **274**, 2761–7.
- 39 Buckling A, Wills M a & Colegrave N (2003) Adaptation limits diversification of experimental bacterial populations. *Science* **302**, 2107–9.
- 40 Spiers AJ, Kahn SG, Bohannon J, Travisano M & Rainey PB (2002) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics* **161**, 33–46.
- 41 Brockhurst M a, Hochberg ME, Bell T & Buckling A (2006) Character displacement promotes cooperation in bacterial biofilms. *Curr. Biol.* **16**, 2030–4.
- 42 Goymer P, Kahn SG, Malone JG, Gehrig SM, Spiers AJ & Rainey PB (2006) Adaptive Divergence in Experimental Populations of *Pseudomonas fluorescens*. II. Role of the GGDEF Regulator, WspR, in Evolution and Development of the Wrinkly Spreader Phenotype. *Genetics* **173**, 515–526.

- 43 McDonald MJ, Gehrig SM, Meintjes PL, Zhang X-X & Rainey PB (2009) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. IV. Genetic constraints guide evolutionary trajectories in a parallel adaptive radiation. *Genetics* **183**, 1041–53.
- 44 McDonald MJ, Cooper TF, Beaumont HJE & Rainey PB (2011) The distribution of fitness effects of new beneficial mutations in *Pseudomonas fluorescens*. *Biol. Lett.* **7**, 98–100.
- 45 Beaumont HJE, Gallie J, Kost C, Ferguson GC & Rainey PB (2009) Experimental evolution of bet hedging. *Nature* **462**, 90–3.
- 46 Bantinaki E, Kassen R, Knight CG, Robinson Z, Spiers AJ & Rainey PB (2007) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. *Genetics* **176**, 441–53.
- 47 Ferguson GC, Bertels F & Rainey PB (2013) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. V. Insight into the niche specialist fuzzy spreader compels revision of the model *Pseudomonas* radiation. *Genetics* **195**, 1319–1335.
- 48 Luria S & Delbrück M (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491–511.
- 49 Hall BM, Ma C-X, Liang P & Singh KK (2009) Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbrück fluctuation analysis. *Bioinformatics* **25**, 1564–5.
- 50 Losos JB, Jackman T, Larson A, Quiroz K & Rodriguez-Schettino L (1998) Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279**, 2115–2118.
- 51 Burbrink FT, Chen X, Myers E a, Brandley MC & Pyron RA (2012) Evidence for determinism in species diversification and contingency in phenotypic evolution during adaptive radiation. *Proc. Biol. Sci.* **279**, 4817–26.
- 52 Zhang X-X & Rainey PB (2007) Construction and validation of a neutrally-marked strain of *Pseudomonas fluorescens* SBW25. *J. Microbiol. Methods* **71**, 78–81.
- 53 Lenski RE, Rose MR, Simpson SC, Tadler SC, Naturalist TA & Dec N (1991) Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2,000 Generations. *Am. Nat.* **138**, 1315–1341.
- 54 Rosche WA & Foster PL (2000) Determining Mutation Rates in Bacterial Populations. *Methods* **17**, 4–17.
- 55 Foster PL (2006) Methods for Determining Spontaneous Mutation Rates. *Methods Enzymol.* **409**, 195–213.

Summary

The origin of biological complexity has intrigued scientists for centuries. The foundation of the modern explanation was developed in the 18th and 19th century and later developed into what is known as the theory of evolution. Evolution is the process in which the genetic composition of organisms in a population changes over time. Random mutations in the DNA of organisms cause genetic changes and the frequency of individuals in a population that harbour such mutations can change due to a range of processes. Some genetic changes cause heritable changes in the phenotypes of organisms. When such phenotypic alteration result in an increased organismal fitness, which can be defined as its ability to survive and reproduce, the frequency of the individuals harbouring the mutations gradually increases in the population. This process is one of the drivers of evolutionary changes and is called natural selection.

Despite many major advances in our understanding of the evolutionary process, many of the underlying mechanisms still remain clouded in a mist of unanswered questions, which is not in the last place due to the incredibly intricate way in which changes in the genetic material result in phenotypic changes. This makes it difficult to understand how the phenotypic variation, which provided the raw material for the evolution of life's incredible diversity and complexity, can have been generated. In addition, the complex interplay between mutation and natural selection further complicates matters. Natural selection emerges from the interactions between organisms and their biotic and abiotic environment. The roles of these, often very complex, interactions are difficult to disentangle from the role of the mutation and the potential of organisms to generate phenotypic variation upon mutation that can facilitate evolution (evolvability). The research described in this thesis aimed to contribute to our understanding of the mechanisms that facilitate evolution, by shedding light on the creative power of a modular mode of mutation (compositional evolution) and isolating the role of natural selection from evolvability during the evolution of biological diversity from a single ancestral type.

In the first experimental chapters of this thesis, we focussed on the mechanisms underpinning the evolution of protein complexes by compositional evolution. During compositional evolution, protein complexes emerge, innovate and adapt by integration of pre-existing proteins that functioned in a different context. Most of what we know about this mode of evolution comes from comparative genomics studies and more experimental insight is necessary to further establish and understand it. In the final experimental chapter we zoomed out and examined the evolution of biodiversity at the level of populations by the process of adaptive radiation. This process has been studied extensively and much is known about the influence of environmental factors. However, the influence of organismal factors—more specifically the influence of the ecology of the ancestor and how it influences natural selection for diversification—is less well understood, despite its importance.

Compositional evolution of protein complexes is defined as evolutionary incorporation of pre-existing parts allowing modification of existing protein complexes or generation of novel ones. While most of what we know about this process comes from comparative genomics studies, experimental evidence is scarce and many questions remain. One requirement for compositional evolution to be

possible is that the genes and corresponding proteins (in this thesis together called *foreign components*) are either immediately compatible with the host complex and other cellular systems or that they can be integrated along selectively accessible mutational trajectories. In order to allow one to think about compositional evolution in a structured manner, we constructed a conceptual framework. Briefly, we assumed a situation in which all protein-encoding genes present in the biosphere could be co-opted for compositional evolution and divided the foreign components into three categories. The first category consists of components that are already compatible and can be modified and functionally integrated by evolution (*evolutionary integration*). The second category entails the foreign components that are incompatible initially but can be evolutionarily integrated. The third category comprises foreign components that are incompatible initially and cannot be modified and functionally integrated by evolution within a defined interval of time. As such, the creative potential of compositional evolution in this conceptual framework is constrained by what may be called a *compatibility horizon*, from which beyond no foreign components can be co-opted for the origin, innovation and adaptation of protein complexes.

In **Chapter 2**, we introduced our model system and examined compositional evolution by systematically investigating the compatibility of foreign, orthologous components from functionally homologous complexes and the capacity of evolution to functionally integrate components in the vicinity of the compatibility horizon. In order to achieve this, we replaced the stator genes in cells of *Escherichia coli* for 22 different stator genes from 15 different bacterial species. The original stator genes encode proteins that are an essential part of the bacterial flagellar motor (BFM), a protein complex that powers bacterial swimming. Replacing these genes for foreign genes resulted in chimeric BFMs (cBFMs) and allowed us to easily examine the compatibility of foreign components by measuring the motility of a population of cells in semi-solid agar. Furthermore, this method also allowed us to study the dynamics of evolutionary integration in the vicinity of the compatibility horizon, by allowing non or poorly motile cells to grow in this environment and thus imposing selection for improved motility. We showed that our model was capable of capturing dynamics of evolutionary integration and that the potential of compositional evolution was possibly limited by several structural and functional factors. Furthermore, our results showed that evolutionary integration of a single foreign component could occur along parallel genetic and phenotypic trajectories. Replicate selection lines of our engineered strains revealed that different, multi-step mutational trajectories were possible for evolutionary integration and that these differences were partly a result of epistatic constraints imposed by earlier mutations.

The mutations underlying foreign component integration were identified by Illumina Next Generation Sequencing in **Chapter 3**. Here we focussed on the overall distribution of the mutations that had accumulated in the lineages and not on individual evolutionary lineages. Because the results in Chapter 2 indirectly showed the existence of mutational differences, we wondered whether we could detect these differences directly both within and between cBFM strains. Also, we were interested in the distribution of mutations. Whole-genome re-sequencing results revealed both within and between cBFM strain differences in mutations found. However, some mutations occurred in multiple cBFM strains. These results indicated that unique mutational steps were accompanied by more general mutational steps and corroborated that a single foreign stator could be incorporated by multiple mutational trajectories. Furthermore, we found that mutations were clustered, especially

around the foreign stator proteins. We proposed that mutations in components in the structural vicinity of the foreign component conferred the largest beneficial effect on motility.

The phenotypic and mutational data described in chapters 2 and 3 did not reveal the detailed mutational order of individual evolutionary lineages. In **Chapter 4** we analysed the temporal pattern of mutations for a number of focal lineages in detail. Also we examined the step-wise phenotypic changes below the population level, including the single-cell swimming speed and the percentage of non-motile cells. We showed that the evolutionary lineages of different and even the same cBFM strain followed different mutational trajectories. All examined evolutionary lineages exhibited different phenotypic dynamics at both population and lower-level phenotypes.

In the chapters we described above, we identified mutations that arose in our selection experiment and also ordered them for a number of focal evolutionary lineages. However, we did not resolve the contribution of individual mutations to integration of the foreign stators. Therefore, in **Chapter 5** we investigated this for a specific selection of mutations: mutations in the stator that facilitated the first adaptive step. With the results, we intended to shed light not only on the diversity of mutations that could lead to compatibility or improved integration, both between cBFM strains and within a single strain, but also on the location in the proteins in which these mutations were located and how these correlated with their functional effects. Our results revealed that first-step compatibilizing mutations in the foreign stators differed both within and between cBFM strains. Furthermore, we showed that different mutations resulted in different motility levels and that mutations were not restricted to a specific part of the foreign components. Finally, we showed that many, but not all, mutations in stator proteins led to a large increase in motility.

Taken together, the results described in the first four experimental chapters of this thesis provide valuable insight into the functional incorporation of orthologous components from functionally similar protein complexes. These observations demonstrate the capacity of evolution to integrate pre-adapted, but sub- or incompatible, components, a process likely to have facilitated the evolution of protein complexes in natural history.

In the final chapter of this thesis, **Chapter 6**, we shifted our attention to the evolution of complexity at a higher level of biological organization: ecological diversity at the community level. Using experimental populations of the soil-bacterium *Pseudomonas fluorescens*, we investigated the influence of organismal factors on the evolution of biodiversity in the process of *adaptive radiation*: the rapid diversification of a single ancestor into organisms with different ecological niches. We started by making a distinction between organismal evolvability, which is the capacity of organisms to generate heritable and selectable phenotypic variation, and the niche of organisms, which is the complex of reciprocal ecological interactions between organisms and the environment. This distinction allowed us to examine the independent roles of these factors in constraining the evolution of genotypes with diverse ecological strategies from a single ancestral type. Our results showed that the ecological properties of the ancestor constrained the rate of adaptive radiation. This finding represents the first empirical support for a prediction of the ecological theory of adaptive radiation: that the niche of the founder has an impact on the dynamics of adaptive radiation.

In the introduction we stated that one of the biggest challenges in biology is to answer the question of how the biological complexity that we witness today came about? In this thesis we set out to generate new experimental insight that helps to answer this question. We investigated

evolution of complexity at two different levels, namely at the level of protein complexes and at the level of populations. At the level of protein complexes we generated new empirical evidence for the creative capacity of compositional evolution, which is a key process in the current evolutionary explanation of the evolution of thousands of protein complexes. In particular, we showed that evolution can overcome incompatibilities between a prospective pre-adapted foreign component and an existing complex. At the level of populations, we managed to separate evolvability and ecological properties of organisms, which allowed us to test the prediction that the niche of the founder of an adaptive radiation impacts the evolutionary dynamics.

The research into compositional evolution of protein complexes described in this thesis combined experimental evolution and synthetic biology in order to investigate the mechanisms behind a key step in evolution. We believe that this approach, and future versions of it that make use of more advanced, high throughput techniques, provide a very powerful way to test and expand evolutionary theory. I am thrilled to find out in the coming years what (molecular) pieces will be added to the enormous, billion years old, evolutionary puzzle.

Samenvatting

Het ontstaan van biologische complexiteit heeft wetenschappers sinds lange tijd geïntrigeerd. De basis voor de huidige verklaringen was ontwikkeld in de 18^e en 19^e eeuw en is later doorontwikkeld in wat bekend staat als de theorie van evolutie. Evolutie is het proces in welke de genetische samenstelling van organismen in een populatie verandert over tijd. Toevallige mutaties in het DNA van organismen zorgen voor genetische veranderingen en de frequentie in de populatie van individuen die die mutaties dragen kan veranderen door verschillende processen. Sommige genetische veranderingen zorgen voor erfelijke veranderingen in de fenotypes van organismen. Wanneer zulke fenotypische veranderingen resulteren in een verhoogde fitness van de organismen, welke gedefinieerd is als de mogelijkheid om te overleven en voor nageslacht te zorgen, dan gaat de frequentie van de individuen die die mutaties dragen langzaam aan omhoog in de populatie. Dit proces is één van de drijvende krachten achter evolutionaire veranderingen en staat bekend als natuurlijke selectie.

Ondanks vele grote vooruitgangen in ons begrip van evolutionaire processen, blijven vele van de onderliggende mechanismen gehuld in een mist van onbeantwoorde vragen. Dit komt niet in de laatste plaats door de ongelooflijk gecompliceerde manieren waarin veranderingen in het genetische materiaal resulteren in fenotypische veranderingen. Dit zorgt ervoor dat het lastig is om te begrijpen hoe fenotypische variatie, welke zorgt voor het ruwe materiaal voor de evolutie van de ongelooflijke diversiteit en complexiteit van het leven op Aarde, gegenereerd kan worden. Daarbij zorgt de complexe interactie tussen mutatie en natuurlijke selectie voor meer onduidelijkheid. Natuurlijke selectie ontstaat uit de interactie tussen organismen en hun biotische en abiotische omgeving. De rollen van deze, vaak uitermate complexe, interacties is lastig los te koppelen van de rol van mutaties en de mogelijkheid van organismen om fenotypische variatie te genereren als gevolg van mutaties die evolutie faciliteren (*evolvability*). Het doel van het onderzoek dat is beschreven in dit proefschrift was om bij te dragen aan ons begrip van de mechanismen die evolutie faciliteren, door middel van het creëren van meer duidelijkheid over de creativiteit van een modulaire manier van mutatie (compositional evolutie) en door middel van het isoleren van de rol van natuurlijke selectie van *evolvability* gedurende de evolutie van biologische diversiteit door een enkele voorouder.

In de eerste experimentele hoofdstukken van dit proefschrift hebben wij ons gefocust op de mechanismen die ten grondslag liggen aan de evolutie van eiwitcomplexen door middel van compositionele evolutie. Tijdens compositionele evolutie ontstaan, innoveren en passen eiwitcomplexen zich aan, door middel van integratie van eiwitten die al bestonden en in een andere context functioneerden. Het merendeel van wat we weten over deze manier van evolutie komt voort uit onderzoeken die zich baseren op *comparative genomics* en meer experimenteel inzicht is noodzakelijk om het beter te begrijpen. In het laatste hoofdstuk hebben wij evolutie van biodiversiteit op het niveau van populaties onderzocht, welke gegenereerd werd door het proces van adaptieve radiatie. Dit proces is al grondig bestudeerd en er is al veel bekend over de invloed van omgevingsfactoren. Er is echter veel minder bekend over de invloed van organisme-specifieke factoren, en meer specifiek de invloed van de ecologie van de voorouder en hoe dat natuurlijke selectie beïnvloedt, ondanks het belang hiervan.

Composicionele evolutie van eiwitcomplexen is gedefinieerd als de evolutionaire incorporatie van eiwitten die al bestaan, wat resulteert in de modificatie van een eiwit complex of het ontstaan van nieuwe complexen. Terwijl het meeste van wat we over deze manier van evolutie weten komt van onderzoeken gebaseerd op *comparative genomics*, is er weinig experimenteel bewijs en veel vragen blijven dan ook onbeantwoord. Eén voorwaarde die composicionele evolutie mogelijk maakt, is dat de genen en corresponderende eiwitten (welke in dit proefschrift gezamenlijk *vreemde componenten* worden genoemd) ofwel direct compatibel zijn met het al bestaande eiwitcomplex en andere cellulaire systemen of dat ze geïntegreerd kunnen worden via mutationale routes die selectief ook mogelijk zijn. Om men gestructureerd te laten nadenken over composicionele evolutie, hebben wij een conceptueel raamwerk bedacht. We hebben aangenomen dat alle genen die voor eiwitten coderen en aanwezig zijn in de biosfeer gebruikt kunnen worden voor composicionele evolutie en hebben alle vreemde componenten onderverdeeld in drie categorieën. De eerste categorie bestaat uit componenten die direct compatibel zijn en direct gemodificeerd en functioneel geïntegreerd kunnen worden door evolutie (dit noemen we evolutionaire integratie). De tweede categorie bestaat uit de vreemde componenten die niet direct compatibel zijn, maar die wel evolutionair geïntegreerd kunnen worden. De derde categorie bestaat uit de vreemde componenten die niet direct compatibel zijn en die ook niet functioneel geïntegreerd kunnen worden door evolutie in een bepaald tijdsinterval. Op deze manier wordt de creativiteit van composicionele evolutie in ons conceptuele raamwerk beperkt door een zogenaamde *compatibiliteitshorizon*, van waarachter geen vreemde componenten gebruikt kunnen worden voor het ontstaan, innoveren en aanpassen van eiwitcomplexen.

In **hoofdstuk 2** hebben wij ons model systeem geïntroduceerd en hebben wij composicionele evolutie onderzocht door systematisch de compatibiliteit van vreemde componenten uit functioneel homologe eiwitcomplexen te controleren en door te kijken of de componenten functioneel geïntegreerd konden worden door evolutie. Om dit voor elkaar te krijgen hebben wij de stator genen in cellen van *Escherichia coli* vervangen door 22 verschillende stator genen uit 15 verschillende soorten bacteriën. De originele stator genen coderen voor eiwitten die een essentieel deel zijn van de Bacteriële Flagellaire Motor (BFM), een eiwit complex dat zorgt voor het zwemmen van bacteriën. Het vervangen van deze genen voor vreemde genen resulteerde in chimere BFMs (cBFMs) en stelde ons in staat om op een gemakkelijke manier de compatibiliteit van de vreemde componenten te controleren, namelijk door middel van het meten van de beweeglijkheid van een populatie van cellen in half-hard agar. Daarbij stelde deze methode ons in staat om de dynamiek van evolutionaire integratie te bestuderen in de buurt van de compatibiliteitshorizon, namelijk door niet of nauwelijks bewegende cellen te laten groeien in deze omgeving en dus een selectie uit te oefenen op verbeterde beweeglijkheid. We hebben laten zien dat ons model system in staat was om de dynamiek van evolutionaire integratie vast te leggen en dat de mogelijkheden van composicionele evolutie mogelijk beperkt werden door verschillende structurele en functionele factoren. Daarnaast hebben onze resultaten ook laten zien dat de evolutionaire integratie van een enkele vreemde component plaats kon vinden langs parallele genetische en fenotypische routes. Replica's van selectielijnen toonden dat verschillende, multi-staps mutationale routes mogelijk waren voor evolutionaire integratie en dat de gevonden verschillen gedeeltelijk verklaard konden worden door epistatische beperkingen die werden opgelegd door eerdere mutaties.

De mutaties die ten grondslag lagen aan de integratie van de vreemde componenten, werden geïdentificeerd door middel van Illumina Next Generation Sequencing in **hoofdstuk 3**. In dit hoofdstuk focusten wij ons op de algemene verdeling van alle mutaties in alle selectielijnen en niet op individuele selectielijnen. Omdat de resultaten van hoofdstuk 2 indirect al aantoonde dat er mutationale verschillen waren, vroegen wij ons af of wij deze verschillen konden vinden zowel binnen als tussen verschillende cBFM stammen. Daarnaast waren wij geïnteresseerd in de verdeling van mutaties. De resultaten van de genetische analyse toonden mutationale verschillen zowel binnen als tussen verschillende cBFM stammen. Sommige mutaties werden echter ook in verschillende cBFM stammen gevonden. Dit toonde aan dat unieke mutationale stappen werden vergezeld door generieke mutationale stappen en bood nogmaals het bewijs dat een enkele vreemde component geïntegreerd kon worden via verschillende mutationale routes. Daarnaast toonden de resultaten dat de mutaties geclusterd waren, voornamelijk rondom de vreemde componenten. Naar aanleiding van de resultaten stelden wij voor dat mutaties in componenten in de structurele buurt van de vreemde componenten resulteerden in de grootste voordelige effecten op beweeglijkheid.

De fenotypische en mutationale data die beschreven werd in hoofdstukken 2 en 3, toonde niets over de mutationale volgorde in individuele selectielijnen. In **hoofdstuk 4** analyseerden wij deze mutationale volgorde voor aan aantal selectielijnen. Ook bestudeerden wij stapsgewijze fenotypische veranderingen op een lager niveau dan dat van de populatie, waaronder de zwemsnelheid van enkele cellen en het percentage niet-beweeglijke cellen. We lieten zien dat de selectielijnen van verschillende, maar ook van dezelfde, cBFM stammen verschillende mutationale routes namen. Alle bestudeerde selectielijnen toonden verschillende fenotypische dynamieken op zowel populatie niveau als op het niveau van een enkele cel.

In de hoofdstukken die we hierboven beschreven hebben, hebben we mutaties geïdentificeerd die in ons selectie-experiment waren ontstaan en ook hebben wij de volgorde bepaald voor een aantal selectielijnen. Wij hebben echter nog niet gekeken naar de invloed van afzonderlijke mutaties op de integratie van de vreemde componenten. Daarom hebben wij dit bestudeerd in **hoofdstuk 5** voor een bepaald type mutaties, namelijk de mutaties in de vreemde componenten die een rol speelden in de eerste adaptieve stap. Met deze resultaten hoopten wij niet alleen iets te laten zien van de diversiteit aan mutaties die kunnen leiden tot compatibiliteit of verbeterde integratie, zowel tussen cBFM stammen als binnen één cBFM stam, maar hoopten wij ook iets te laten zien van de locatie van de mutaties in de eiwitten en hoe dit correleerde met hun functionele effecten. Onze resultaten toonden dat de mutaties in de vreemde componenten in de eerste stap die resulteerden in compatibiliteit, verschilden zowel tussen als binnen cBFM stammen. Bovendien lieten wij zien dat verschillende mutaties resulteerden in verschillende niveaus van beweeglijkheid en dat mutaties niet beperkt waren tot een specifiek gedeelte van de vreemde componenten. Tot slot toonden onze resultaten dat veel mutaties in de vreemde componenten, maar niet alle, resulteerden in een grote toename in beweeglijkheid.

Samenvattend bieden de resultaten die wij beschreven hebben in de eerste vier hoofdstukken van dit proefschrift waardevolle nieuwe inzichten in de functionele incorporatie van orthologe componenten uit functioneel gelijkwaardige eiwitcomplexen. De observaties demonstreerden dat evolutie in staat is om bestaande, sub- of incompatibele, componenten functioneel te integreren,

een proces dat de evolutie van eiwitcomplexen in de natuurlijke geschiedenis gefaciliteerd kan hebben.

In het laatste hoofdstuk van dit proefschrift, **hoofdstuk 6**, hebben wij onze focus verlegd naar de evolutie van complexiteit op een hoger niveau van biologische organisatie: ecologische diversiteit op het niveau van gemeenschappen. Gebruik makend van experimentele populaties van de groundbacterie *Pseudomonas fluorescens*, hebben wij de invloed van verschillende factoren van organismen op de evolutie van biodiversiteit bestudeerd tijdens het proces van *adaptieve radiatie*: de snelle diversificatie van een enkele voorouder in organismen met verschillende ecologische niches. Om te beginnen hebben wij onderscheid gemaakt tussen *evolvability* van organismen, wat de capaciteit van organismen is om erfelijke en selecteerbare fenotypische variatie te genereren, en de niche van organismen, wat bestaat uit het geheel van reciproke ecologische interacties tussen organismen en de omgeving. Het gemaakte onderscheid stelde ons in staat om de onafhankelijke rol van deze factoren te bestuderen in het beïnvloeden van de evolutie van genotypen met diverse ecologische strategieën uit één enkele voorouder. Onze resultaten lieten zien dat de ecologische eigenschappen van de voorouder van invloed waren op de snelheid van adaptieve radiatie. Dit resultaat is het eerste empirische bewijs voor een voorspelling gedaan in de ecologische theorie van adaptieve radiatie, namelijk dat de niche van het organisme aan de basis van de adaptieve radiatie invloed heeft op de dynamiek ervan.

In de introductie stelden we dat één van de grootste uitdagingen in de biologie is om antwoord te vinden op de vraag hoe de huidige biologische complexiteit is ontstaan. In dit proefschrift hebben wij getracht om nieuw, experimenteel inzicht te genereren dat kan helpen om deze vraag te beantwoorden. We hebben de evolutie van complexiteit op twee verschillende niveaus bestudeerd, namelijk op het niveau van eiwitcomplexen en op het niveau van populaties. Op het niveau van eiwitcomplexen hebben wij empirisch bewijs geleverd voor de creativiteit van compositionele evolutie, wat een belangrijk proces is in de huidige verklaring voor de evolutie van duizenden eiwitcomplexen. Wij hebben laten zien dat evolutie incompatibiliteiten tussen vreemde componenten en een al bestaand eiwitcomplex kan opheffen. Op het niveau van populaties zijn we erin geslaagd om een onderscheid te maken tussen *evolvability* en ecologische eigenschappen van organismen, wat ons in staat stelde om de hypothese te testen dat de niche van een organisme dat aan de basis staat van een adaptieve radiatie van invloed is op de evolutionaire dynamiek.

In het onderzoek naar compositionele evolutie van eiwitcomplexen dat is beschreven in dit proefschrift, hebben wij gebruik gemaakt van experimentele evolutie en synthetische biologie. Wij geloven dat deze aanpak, en ook toekomstige versies die gebruik maken van doorontwikkelde, *high throughput* technieken, een zeer krachtige manier is om evolutionaire theorie te testen en uit te breiden. Om deze reden kijken wij dan ook erg uit naar welke (moleculaire) stukken aan de enorme, miljarden jaren oude, puzzel toegevoegd gaan worden in de komende jaren.

Dankwoord

Toen ik in september 2010 begon aan mijn promotietraject, had ik niet kunnen voorspellen hoe het zou verlopen. Bijna zeven jaar later kijk ik met gemengde gevoelens terug op de periode als promovendus, maar ben ik wel enorm blij dat ik het uiteindelijk afgerond heb. Ik heb er veel van geleerd en ongetwijfeld zal deze waardevolle ervaring nog van pas komen.

Allereerst wil ik mijn co-promotor Bertus bedanken voor de mogelijkheid om na mijn afstuderen het onderzoek voort te zetten, uit te breiden en ook andere onderwerpen te mogen ontdekken. Tevens was dit een goede mogelijkheid om mij verder te ontwikkelen als onderzoeker. Ik ben ervan overtuigd dat jouw enthousiasme, expertise, inzicht en doorzettingsvermogen zullen leiden tot mooie ontdekkingen op het gebied van evolutionaire biologie en ik zal hier ook graag nog over willen lezen. Ook wil ik je bedanken voor jouw inspanningen om het proefschrift af te ronden. We hadden beiden een andere kijk op hoe het er nu uit moest komen te zien, wat er aan heeft bijgedragen dat het proces van schrijven *iets* meer tijd in beslag nam dan wat we van te voren hadden bedacht. Maar we zijn doorgedaan. Met ups en downs hebben we ons hier doorheen geslagen en nu is het boekje klaar! Bedankt voor alles, uiteraard ook de aangename gesprekken en discussies over andere onderwerpen dan de werkgerelateerde.

Erwin, bedankt voor jouw inzet in het lab en de vele keren dat je iets voor me hebt uitgezocht, aangeënt, ingevroren of wat dan ook. En niet te vergeten het ritje naar het ziekenhuis. Bij het binnenkomen van de Biobrick was het direct duidelijk of je aanwezig was of niet: jouw lach, stem en gezang reiken ver! Jouw vaak vrolijke houding maakte het labwerk niet alleen leuk, maar hebben mij ook geleerd om geluiden af te schermen wanneer concentratie nodig is ☺. Wanneer je een 96-wells plaat moet vol pipetteren met verschillende stoffen en in verschillende houdingen, dan kan een zingend persoon soms best afleidend zijn. Maar ga er vooral mee door, want het was altijd gezellig!

Bert en Thierry, ook jullie wil ik bedanken voor jullie expertise, hulp en adviezen. Ik heb het altijd als zeer fijn ervaren om met jullie te werken. Millie, we might have had our differences, but also I want to thank you for your input, work and knowledge. I think we managed to obtain some interesting results! Carsten en Esmee, de resultaten van jullie werk zijn niet opgenomen in dit proefschrift, maar toch wil ik jullie bedanken voor al jullie harde werk. Het uitpluizen van al de ecologische strategieën was niet niets en leverde mooie inzichten op. Ook gaat mijn dank uit naar Max, Stefanie, Marcel en Steven voor hun werkzaamheden in het BFM project. Om verder niemand te vergeten, wil ik tevens iedereen bij Bionanoscience bedanken voor alle gesprekken en discussies.

In juni 2015 ben ik aan de slag gegaan bij het Instituut Fysieke Veiligheid. Op dat moment was mijn proefschrift nog niet afgerond. Bij het IFV ben ik terecht gekomen in een voor mij enorm stimulerende omgeving; ik wil dan ook iedereen bedanken die heeft bijgedragen aan mijn tijd als trainee. Specifiek wil ik bedanken Lydia en Wim voor de mogelijkheid om aan mijn proefschrift te werken tijdens werktijd. Tevens wil ik Nils, Menno en Ricardo bedanken voor de gesprekken die mij meer motivatie gaven om het proefschrift af te ronden.

Als laatste, maar zeker niet als minste, wil ik mijn vrouw Diana bedanken. Al jaren ben je voor mij een steun en toeverlaat en heb jij enorm veel klaagpartijen moeten aanhoren. Toch wist jij het altijd weer positief te bekijken en mijn motivatie aan te wakkeren. Ook weet je altijd de juiste vragen te

stellen om mij weer op gang te helpen. Onze lieve dochter Irsa heeft ons leven er niet rustiger op gemaakt, maar zeker wel enorm veel meer waarde gegeven. Het schrijven van mijn proefschrift kreeg hierdoor uiteraard minder aandacht. Toch heb je me vaak de mogelijkheid gegeven om nog even aan het proefschrift te werken, terwijl jij zorgde voor onze kleine deugniet. Hoe eerder af, hoe eerder we meer tijd met elkaar konden doorbrengen. Inmiddels is ons tweede kindje op komst, dus het is maar goed dat het moment nu eindelijk is aangebroken: het proefschrift is af! Ik kijk ernaar uit om op nog vele en vele zondagen niet meer te hoeven schrijven, maar om lekker met jullie te zijn. Ik hou van jullie!

Curriculum Vitae

Régis Christian Emil Flohr

February 16, 1985	Born in Leiderdorp, The Netherlands
1997-2003	Secondary education Stedelijk Gymnasium Leiden
2004-2007	B.Sc. Life Science and Technology Leiden University and Delft University of Technology
2007-2010	M.Sc. Life Science and Technology Leiden University and Delft University of Technology
2010-2014	Ph.D. research under supervision of dr. Hubertus J.E. Beaumont Delft University of Technology

List of publications

Régis C.E. Flohr, Magdalena D. Pieczynska, Thierry K.S. Janssens, Min-Young Heo, Erwin van Rijn, Jeremie Capoulade, Francesco Pedaci & Hubertus J.E. Beaumont (2017) Compositional evolution of the bacterial flagellar motor. *Manuscript in preparation*.

The manuscript is based on Chapters 2-5 and data from additional experiments carried out in the Beaumont lab. This additional work (i) quantified foreign stator expression levels, (ii) demonstrated transient epistatic constraints on their evolutionary integration engendered by early adaptive stator mutations and (iii) showed that foreign stators 8, 13, 29 resulted in non functional BFM (no filament rotation detectable).

Régis C.E. Flohr, Carsten J. Blom, Paul B. Rainey and Hubertus J.E. Beaumont (2013) Founder niche constrains evolutionary adaptive radiation, *PNAS*, 110 (51), 20663-20668