Diversity Generation in Evolving Microbial Populations

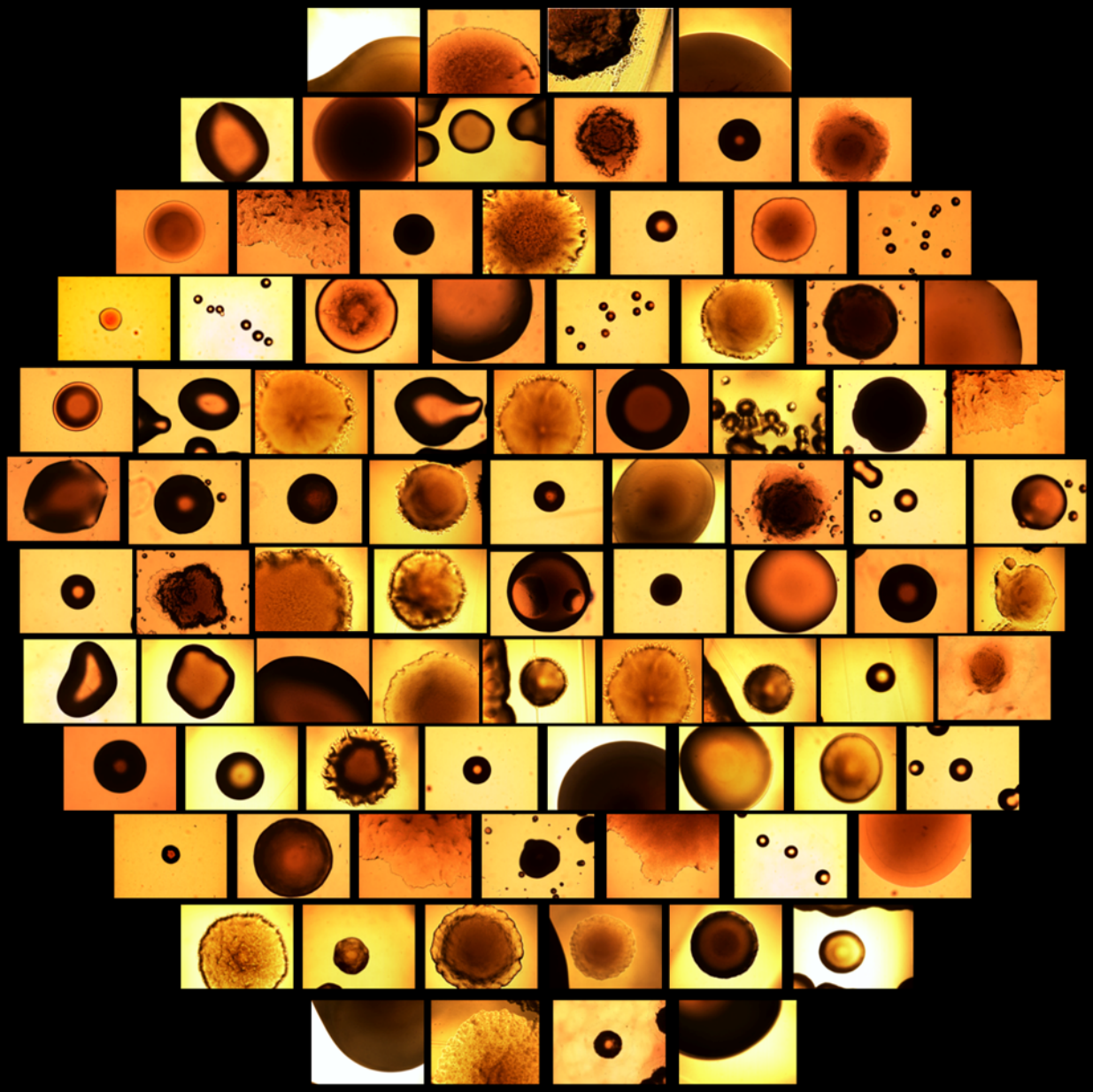
Markussen, Trine

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

DIVERSITY GENERATION IN EVOLVING MICROBIAL POPULATIONS

DTU



PhD Thesis

By

Trine Marenda Markussen

DEPARTMENT OF SYSTEMS BIOLOGY

Preface

This thesis is submitted as a partial fulfilment of the requirements to obtain a PhD degree at the Technical University of Denmark (DTU). The work presented here was carried out from January 2011 to January 2014 at the Infection Microbiology Group (IMG), Department of Systems Biology at DTU under the supervision of Associate Professor Lars Jelsbak and Professor Søren Molin. The work was co-funded by DTU and Danish Research Council.

Trine Marenda Markussen

Vanløse, January 2014

Acknowledgements

"Life is nothing without a little chaos to make it interesting"

- Amelia Atwater-Rhodes

I wish to express my gratitude to a lot of people who supported and helped me during my PhD. First of all, I want to thank Lars Jelsbak for his great guidance and for being an inspiring supervisor, who always find the time for my questions and for listening to all my struggles. I also wish to thank my co-supervisor Søren Molin for first introducing me to the exiting area of bacterial evolution and adaptation to novel environments and for his excellent supervision and boundless expertise.

A special thank to the 'sputum meeting' members and collaborators: Helle Krog Johansen, Niels Højby, Kasper Aanæs, and Oana Ciofu for excellent collaboration, good discussions and helpful comments at our sputum meetings.

Thank you, Rasmus L. Marvig for teaching me about phylogenetic trees and other bioinformatics questions. While my projects mainly focused on the DK1 clone, Rasmus worked with the DK2 and we therefore have had many discussions on our results. Thanks to Vinoth Wigneswaran for all the good weekends we shared in the lab, and our Monday morning discussions on how to make the world a better place.

Especially, I wish to thank the past and present members of the IMG group for creating a great atmosphere, lots of good time (not only in the lab), scientific discussions and your helpful skill – thank you very much Lea M. Madsen, Eva K. Andresen, Søren Damkiær, Nicholas Jochumsen, Linda R. Jensen, Maria G. Lozano, Susanne (Søs) Koefoed, Alexandra E. Burleigh, Charlotte F. Michelsen, Juliane T. Thøgersen, Fatima Yousef, Lei Yang, Claus Sternberg, Cristina I. A. Hierro, Hossein Khademi and all the many students and visitors of the group. A huge thanks to the technical and administrative staff; Lisser St. Clair-Norton, Lone Hansen, and Janina Brøker.

I wish to express my gratitude to my Canadian colleagues from the amazing lab of Joesph Lam. It was truly a pleasure to work with you, eh, and I never forget your hospitality. A special thanks to Joe himself for taking me in as one of his own and for giving me an insight into the world of LPS.

Last but definitely not least, I want to thank my family for their endless support, patience, and understanding. A special thank to my dear Rasmus K. Bojsen (soon-to-be Markussen) for loving and supporting me when I was impossible.

Abstract

Organisms have evolved and diversified since the beginning of life. Although, generation and maintenance of diversity within ecosystems has been a central concern in ecology and evolutionary biology, little is known of the evolutionary processes driving diversification. Especially, diversification in relation to chronic infection is a major concern as high population diversity has been predicted to result in survival and persistence of the infecting microbe.

Therefore, understanding within-host dynamics and population diversification is necessary for optimal diagnosis and therapeutic treatment.

Chronic *Pseudomonas aeruginosa* infections in the airways of patients with cystic fibrosis (CF) offer opportunities to study bacterial evolution and adaptation in natural environments. Significantly phenotypic and genomic changes of *P. aeruginosa* have been observed during chronic infection. While *P. aeruginosa* diversity has been documented in contemporary respiratory specimens, it is less clear to what extent within-patient diversity contributes to the overall population structure and whether the population is geographically or homogeneously distributed throughout the airways.

The focus of this thesis has been to get a better understanding of how bacterial populations adapt to new, complex and heterogeneous environments with multiple selective pressures over long periods, and to analyse diversification during this adaptation. Using the *P. aeruginosa* chronic infection as a model system, and by combining bacterial genome sequencing, phenotypic profiling and unique sampling materials which included clonal bacterial isolates sampled for more than 4 decades from chronically infected CF patients, we were able to investigate the diversity generation of the clinical important and highly successful *P. aeruginosa* DK1 clone type during chronic airway infection in CF patients.

We show here that diversification of *P. aeruginosa* DK1 occurs through the emergence of coexisting subpopulations with distinct phenotypic and genomic features and demonstrate that this diversification was a result of niche specialization as each subpopulation colonized separate geographical niches. This highly complex population diversity was observed to be stably maintained during long-term evolution.

Before diversification of the DK1 clone, a regulatory mutation was found to be fixed in the population causing alteration of multiple phenotypes representing the chronic stage phenotype.

Often chronic CF infections are polyclonal and therefore we investigated the population dynamics in a patient polyclonal infected with both DK1 and DK2. We demonstrated that diversification was affected by the presence of other clones; interaction between the two clones resulted in horizontal DNA transfer that contributed to the observed population diversity by creating a novel strain DK1/2 found to persist in the CF airways.

These data indicate that spatial compartmentalization and transfer of DNA between infecting microbes can cause generation and maintenance of population diversity of infecting pathogens. Furthermore, fine-tuning of global regulatory networks by modification of transcriptional regulators has fundamental roles in successful adaptation of *P. aeruginosa* to the CF environment.

Dansk Resumé

Organismer har udviklet sig og diversificeret siden livets begyndelse. Selvom generering og vedligeholdelse af diversitet i økosystemer har været et centralt anliggende i økologi og evolutionær biologi, vides der kun lidt om de evolutionære processer der ligger til grund for diversificering. Diversitet er især et stort problem i forhold til kronisk infektion hvor stor populationsdiversitet kan resultere i overlevelse og persistens af den inficerende mikrobe.

Det er derfor vigtig at forstå dynamikken of populationsdiversiteten i patienten for optimal diagnose og terapeutisk behandling.

Kroniske luftvejsinfektioner med *Pseudomonas aeruginosa* hos patienter med cystisk fibrose (CF) giver muligheden for at studere bakteriel evolution og tilpasning i naturlige miljøer. Signifikante fænotypiske og genomiske ændringer af *P. aeruginosa* er blevet observeret under kronisk infektion. Mens *P. aeruginosa* diversitet er blevet dokumenteret i sputum prøver, er det mindre klart, i hvilket omfang diversiteten bidrager til den samlede populations struktur i patienten, og om populationen er geografisk eller homogent fordelt i luftvejene.

Fokusset i denne afhandling har været at få en bedre forståelse af, hvordan bakteriepopulationer tilpasse sig nye, komplekse og heterogene miljøer med forskellige selektive pres over lange perioder, og at analysere diversiteten i løbet af denne tilpasning. Ved bruge af *P. aeruginosa* kroniske infektioner som modelsystem, og ved at kombinere bakteriel genomsekventering, fænotypiske profiler og unikke prøver, som inkluderer klonale bakterieprøver isoleret fra kronisk inficerede CF-patienter for mere end 4 årtier, var vi i stand til at undersøge diversitet udviklingen af den klinisk vigtige og yderst succesfulde *P. aeruginosa* DK1 klon i kroniske luftvejsinfektioner hos CF-patienter.

Vi viser her, at diversificering af *P. aeruginosa* DK1 sker gennem fremkomsten af sameksisterende subpopulationer med forskellige fænotypiske og genomiske funktioner og viser, at denne diversitet var et resultat af niche specialisering hvor hver subpopulation koloniserede adskilte geografiske nicher. Denne meget komplekse populationsdiversitet viste sig at være stabil under langvarig evolution. Før diversificering af DK1 klonen fandt sted, blev en regulatorisk mutation fikseret i populationen som forårsagede ændringen til en fænotype der repræsenterede den kroniske fase fænotype. Kroniske CF infektioner er ofte polyklonale og derfor undersøgte vi populationsdynamikken hos en patient inficeret med både DK1 og DK2. Vi viste, at diversificering var påvirket af tilstedeværelsen af andre kloner; samspillet mellem de to kloner resulterede i horisontal DNA overførsel, der bidrog til den observerede populationsdiversitet ved at skabe en ny stamme DK1/2 der vareredе ved i CF luftvejene. Disse data indikerer, at rumlig opdeling og overførsel af DNA mellem de inficerende mikroorganismer kan forårsage generering og vedligeholdelse af populationsdiversiteten af patogener. Derudover har finjustering af de globale regulatoriske netværk ved ændring af transskriptionsregulatorer en grundlæggende rolle i en vellykket tilpasning af P. aeruginosa til CF miljø.

Table of Contents

PREFACE I
ACKNOWLEDGEMENTS II
ABSTRACT III
DANSK RESUMÉ V
PUBLICATION LIST VII
ABBREVIATIONS VIII
LIST OF FIGURES IX

CHAPTER 1 1

INTRODUCTION 1
1.1|OUTLINE OF THESIS 1

CHAPTER 2 2

BACTERIAL ADAPTATION AND EVOLUTION 2
2.1|DYNAMICS OF EVOLUTIONARY ADAPTATION 2
2.1.1|ADAPTIVE AND NONADAPTIVE MUTATIONS 3
2.1.2|ADAPTIVE RADIATION AND NICHE SPECIALIZATION 5
2.1.3|REWIRING OF REGULATORY NETWORKS 6
2.1.4|HYPERMUTATORS AND THEIR ROLE IN ADAPTATION 8
2.2|HORIZONTAL GENE TRANSFER IN BACTERIAL EVOLUTION 8

CHAPTER 3 11

EVOLUTION IN NATURAL ENVIRONMENTS 11
3.1|CYSTIC FIBROSIS 11
3.1.1|COMPARTMENTS OF THE CF AIRWAY 12
3.1.2|CF AIRWAY ECOLOGY 13
3.1.3|INFECTION STAGES 15
3.2|PSEUDOMONAS AERUGINOSA 15
3.2.1|CELL-TO-CELL COMMUNICATION IN PATHOGENESIS OF *P. AERUGINOSA* 16
3.2.2|LIPOPOLYSACCHARIDES AND PYOCINS 18
3.3|*P. AERUGINOSA* LIFESTYLE AND ADAPTATION IN CF AIRWAYS 20
3.3.1|PHENOTYPIC ADAPTATION 20
3.3.2|GENOMIC ADAPTATION OF *P. AERUGINOSA* 22

CHAPTER 4 24

PRESENT INVESTIGATIONS 24
4.1|BACKGROUND 24
4.2|THESIS OBJECTIVES AND QUESTIONS 25
4.2|OUTLINE OF STUDIES 25

CHAPTER 5 28

CONCLUSION AND FUTURE PERSPECTIVE 28
5.1|DIFFERENT EVOLUTIONARY PATHS OF DK1 AND DK2 29
5.2|PERSPECTIVES IN RELATION TO EVOLUTIONARY MICROBIOLOGY 31
5.3|MEDICAL PERSPECTIVE 32

REFERENCES 33

CHAPTER 6 41

RESEARCH PAPERS 41

Publication list

Markussen, T. M., Marvig, R. L., Gómez-Lozano, M., Burleigh, A. E., Høiby, N., Johansen, H. K., Molin, S., and Jelsbak, J. (2014) Niche partitioning results in diversification and long-term co-existence of distinct *Pseudomonas aeruginosa* sub-lineages during chronic CF airway infections. *Manuscript submitted.*

Markussen, T. M. M., Nielsen, T., Molin, S., and Jelsbak, L. (2014) Evolutionary modification of *Pseudomonas aeruginosa* Vfr functionality during chronic infection. *Manuscript submitted.*

Markussen, T. M. M., Marvig, R. L., Johansen, H. K., Lam, J. S., and Jelsbak, L. (2014) Generation of genetic diversity among Pseudomonas aeruginosa via horizontal gene transfer during chronic CF infections. *Manuscript in preparation.*

Not included in this thesis

Yang, L., Liu, Y., **Markussen, T.**, Høiby, N., Tolker-Nielsen, T., and Molin, S. (2011) Pattern differentiation in co-culture biofilms formed by Staphylococcus aureus and Pseudomonas aeruginosa. *FEMS Immunol Med Microbiol*, *62(3)*: 339-47

Marvig, R. L., Damkiær, S., Khademi, S. H., **Markussen, T.**, Molin, S., and Jelsbak, J. (2014) Within-Host Evolution of *Pseudomonas aeruginosa* Reveals Adaptation Towards Iron Acquisition from Hemoglobin. *Manuscript in preparation*.

Abbreviations

AHL	*N*-acyl-L-homoserine lactone
ATP	Adenosine triphosphate
BAL	Bronchoalveolar lavage
cAMP	cyclic adenosine monophosphate
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane regulator
CRP	cAMP receptor protein
DNA	Deoxyribonucleic acid
eDNA	Extracellular DNA
ENaC	Epithelial Na^+ channel
HGT	Horizontal gene transfer
LPS	Lipopolysaccharide
LTEE	Long-term evolutionary experiment
PMNs	Polymorphonuclear leukocytes
QS	Quorum sensing
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SNP	Single nucleotide polymorphism
Vfr	Virulence factor regulator

List of figures

Figure 1 Dynamics in an evolving bacterial population.

Figure 2 Climbing adaptive peaks in a rugged fitness landscape.

Figure 3 Trade-offs and specificity of adaptation of genotype A, B, and C.

Figure 4 Adaptive radiation in microcosms.

Figure 5 Evolution of regulatory networks.

Figure 6 Illustration of evolution of four species.

Figure 7 The airway of CF patients.

Figure 8 Prevalence of common human respiratory pathogens in patients with cystic fibrosis as a function of age.

Figure 9 Schematic overview of virulence factors of *P. aeruginosa*.

Figure 10 A simplified overview of the quorum sensing regulation in *P. aeruginosa*.

Figure 11 Structures of LPS in *P. aeruginosa*.

Figure 12 Overview of the sugar residues involved in R-pyocin recognition.

Figure 13 Population heterogeneity of *P. Aeruginosa* within CF airways.

Figure 14 Models of within-patient evolution

Figure 15 Simplified overview of the epidemiology of the DK1 and DK2 clones of *P. aeruginosa* in Danish CF patients attending the Copenhagen CF Centre.

Figure 16 Geographical location of three distinct subpopulations of *P. aeruginosa* in the CF airways.

Figure 17 Evolutionary trajectories of *P. aeruginosa* DK1 clone.

Chapter 1

Introduction

Since the beginning of life, organisms have evolved and diversified. The biological diversity on our planet is one of the most fascinating features we can address as scientists. Fossils or frozen organisms have made it possible to go back in time and study evolutionary processes and trajectories.

In the case of bacteria, we are able to observe 'real-time' evolution as it progress, because of their short generation times and large population sizes in small spaces. Furthermore, transcriptomics, metabolomics, proteomics and next generation sequencing provide great tools for investigating the evolutionary trajectories and dynamics of an adapting population to a novel environment.

Understanding the mechanisms and factors that promote and maintain genetic diversity within ecosystems remains a major topic in ecology and evolutionary biology.

The central focus point in this thesis has been to document how bacterial populations adapt to new, complex and heterogeneous environments over long time periods, and to investigate the occurrence and maintenance of diversity in the evolving populations.

To address this, we used long-term chronic airway infection with *Pseudomonas aeruginosa* of cystic fibrosis (CF) patients for our model system. In this model system, we were able to investigate longitudinal evolutionary trajectories covering more than 150,000 bacterial generations [1] for different bacterial clones in distinct patients.

Knowledge about evolutionary processes and dynamics is important in the treatment of chronic infections. Today, we only have limited information about long-term evolutionary trajectories of pathogens in natural environments.

1.1|Outline of Thesis

This thesis describes the evolutionary dynamics of adapting *P. aeruginosa* to natural environments that are characterized by complexity, spatial structures and dynamic that changes over time. While *Chapter 1* gives an outline of this thesis, *Chapter 2* summarizes some of the important knowledge about bacterial adaptation to novel environment based on *in vitro* evolutionary experiments. *Chapter 3* gives an introduction to the model system used in this thesis: The chronic airway infection of cystic fibrosis patients and the lifestyle of *P. aeruginosa* in the CF airway. Present investigations are introduced in *Chapter 4* that gives the background and main objectives for this thesis. The conclusion and future perspective can be found in *Chapter 5* and the papers presented in this thesis can be found in *Chapter 6*.

Chapter 2

Bacterial adaptation and evolution

For billions of years microorganisms have been adapting and evolving to many fluctuating and hostile environments and have practically colonized all available ecological niches on Earth. Experimental microbial evolution has provided great insight into the microbial dynamics of evolutionary adaptation by allowing researchers to test their theories directly.

Microbes are ideal subjects for observe 'real-time' evolution as they reproduce quickly due to their short generation times and allow large population sizes in small places [2]. Their ability to remain viable at frozen temperatures and resume proliferation when thawed provides the opportunity to compare ancestral and evolved populations.

The experimental setup is mostly simple involving propagating populations from single clones in a defined laboratory environment. After growing for multiple generations, genotypic and phenotypic changes can be compared between ancestral and evolved populations. These kinds of evolution experiments have contributed with knowledge about the dynamics and genetic basis of adaptive processes during evolution.

Evolution experiments with microorganisms have increased significantly during the recent years. In 1988 Richard Lenski and co-workers initiated a long-term evolutionary experiment (LTEE) with *Escherichia coli*, where 12 parallel cultures were inoculated in a simple defined medium. Every day for the next 26 years (and still on-going) the cultures were diluted and re-inoculated in fresh medium to evolve more than 50,000 generation in a constant environment. During this period, samples extracted from the cultures have been stored regularly creating a fossil collection. This collection has provided great insight into the long-term dynamics of adaptation including fitness kinetics and fitness trajectories [3]. Furthermore, impacts of increased mutation rates [4], parallel evolution [5], evolution of innovative traits [6], and the importance of pleiotropic mutations [7] was demonstrated. In this chapter, some of these findings and theories behind adaptive evolution are presented.

2.1|Dynamics of evolutionary adaptation

Most often, an organism needs to change its phenotype in order to adapt to a new environment. This evolutionary process where the reproductive success is improved has for long been of great interest to scientists. An organism may improve its fitness by either phenotypic acclimation or genetic alterations. Phenotypic acclimation refers to the change of phenotype without genetic changes [8]. However, the metabolic cost linked to this may be unfavourable for long-term adaptation to a changing and complex environment [9].

Changing of genetic content can be achieved by either *de novo* mutations in already existing DNA or acquisition of new DNA from other organisms (see section 2.2).

2.1.1|Adaptive and nonadaptive mutations

Adaptation through natural selection may improve the fitness relative to its ancestor by acquisition of mutations including single nucleotide polymorphisms (SNPs), indels (deletions and insertions), and gene rearrangements [2]. Mutations can be divided into beneficial, neutral and deleterious based on the effect on the fitness. Beneficial mutations that drive adaptation by conferring an improved fitness in the environment are also referred to as adaptive mutations and are believed to be caused by the selective pressures rather than being random [10]. Studying adaptive mutations can give better understanding of genetic events underlying the adaptation and can identify genes relevant for this adaptation.

Adaptive benefits of a mutation are usually analysed by transferring the exact mutation to an isogenic strain and investigate fitness effects. [2]. This is necessary since neutral or even deleterious mutations can *hitchhike* with beneficial mutations. Hitchhiking refers to the movement and increase of a neutral or a deleterious mutation in the population caused by its physical linkage with beneficial mutations elsewhere in the genome [2].

Some mutations can appear nonadaptive meaning that they do not have a fitness effect. There may be several reasons for this observation:

1) The mutations are neutral and either hitchhike to high frequency with beneficial mutations or drift randomly to high frequency [2, 11].
2) The mutations are adaptive, however, the methods to determine the fitness are not sensitive enough or not performed in the right conditions to detect the fitness increase [11].
3) The fitness effects of mutations are *epistatic*, meaning that they are modified by their interactions with other mutations in the genome [11, 12].

In this way, some adaptive mutations can appear nonadaptive, but can increase fitness in combination with others.

Even though neutral or deleterious mutations often occur, they will rarely become dominant, while beneficial mutations confer a fitness advantage in the given environment and may therefore be fixed in the population [13]. If the rate of beneficial mutation appearance is low and the fitness advantage of each mutation is large, a beneficial mutation can completely replace all less fit variants and become fixed in the population, a phenomenon known as *selective sweep* [14, 15] (Figure 1). However, if the beneficial mutation rate is high and multiple lineages exist, each with different beneficial mutations, they will compete with each other, an effect known as *clonal interference* [15]. The competition between these different lineages may slow down the rate of fixation, meaning that clonal interference can extent the time of genomic adaptation [16]. If the environment remains constant, one lineage will eventually outcompete the others, a phenomenon called *niche exclusion* [17] (Figure 1).

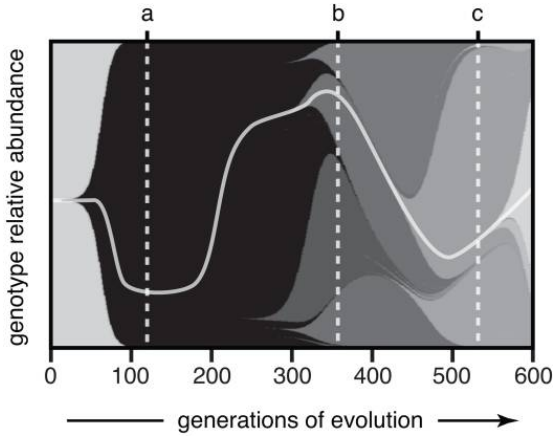


Figure 1|**Dynamics in an evolving bacterial population.** Lineages with new beneficial mutations are represented as shaded wedges that originate in a previous genetic background and rise in frequency as they outcompete their ancestor and other lineages. The same shading indicates the presence of multiple competing beneficial mutations (clonal interference) and the light gray curve highlights the path to the final dominant genotype containing five mutations. (a) A beneficial mutation sweeps to fixation and generates a little diversity. Four lineages with different mutations coexist (b) before the offspring of one lineage become the majority (c) (niche exclusion). Figure is adapted from Barrick and Lenski (2009) [14].

Experimental studies have shown that fitness gains initially are rapid but tend to decelerate over time [18]. This may occur if mutations with large effects become fixed early and/or beneficial mutations become rarer as adaptation progress or have smaller effects on the fitness [2, 18]. This means that a population in a new environment evolves from a region of low fitness towards an adaptive peak [2], which can be achieved by several pathways depending on the fitness landscape and the acquirements of adaptive mutations (Figure 2). When a population is near an adaptive peak only beneficial mutations will move it uphill, while deleterious mutations may move it downhill [19]. When located at the top, the population is referred to as a specialist of the given niche.

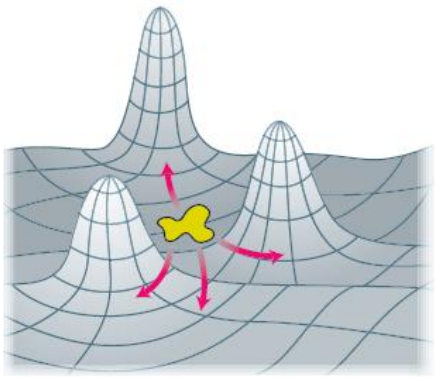


Figure 2|**Climbing adaptive peaks in a rugged fitness landscape.** Demonstration of a population (yellow) evolving in a heterogeneous environment with three distinct niches (peaks). Peaks represent optimal genotype for each niche and arrows represent different genotypic step that can lead to the top. The population may acquire distinct mutations moving it towards different peaks and therefore create diversity. Figure is modified from Elena and Sanjuan (2003) [19].

What happens if the landscape changes or the population is moved to a new environment? It is a general concern in evolutionary biology that beneficial mutations conferring a fitness advantage in one environment, are presumed to have fitness cost in another environment [2, 20]. Such trade-offs are

often caused by *antagonistic pleiotropy*, in which a mutation that increase fitness in one environment, carry a cost and regress of fitness in a second environment. It can also be caused by *accumulation of mutations,* in which neutral mutations in one environment are deleterious in a second environment [2, 20, 21]. In this way, the same mutation that confers adaptation in one environment causes fitness decrease in different environment (Figure 3).

In contrast, generalists can tolerate a broad range of environmental conditions and are expected to arise in changing and variable environments because selection favours the genotype with the highest fitness across all environments [19, 21].

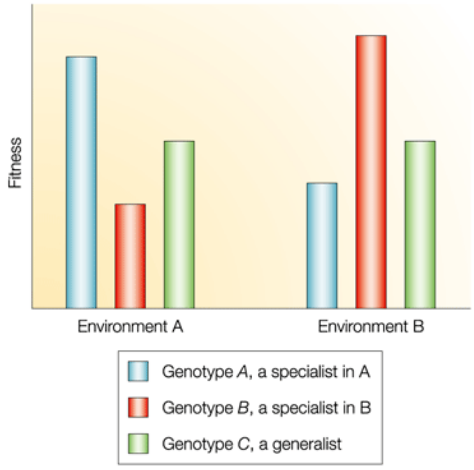


Figure 3|**Trade-offs and specificity of adaptation of genotype A, B, and C.** Genotype A is a specialist adapted to environment A, but have acquired mutations with trade-offs and therefore have reduced fitness in environment B. The opposite is true for genotype B, while genotype C is a generalist that performs moderately well in both environment A and B, but have lower fitness than the specialists in their preferred environment. The figure is adapted from Elena and Lenski (2003) [2].

2.1.2|Adaptive radiation and Niche specialization

In spatially structured environments with distinct adaptive peaks (Figure 2) acquisition of different mutations can generate genomic and phenotypic variation by moving the population towards different adaptive peaks. In doing so, subpopulation may occur that each climb different adaptive peaks in the fitness landscape leading to niche specialisation [19, 22]. This diversification, in which a genetically uniform population evolves into distinct subpopulations, each colonizing their own niches, is also referred to as adaptive radiation.

Rainey and Travisano demonstrated that adaptive radiation increases as a function of environmental complexity and spatial heterogeneity [23] (Figure 4). It was further showed that competition among the evolved niche-specialist maintained variation [24].

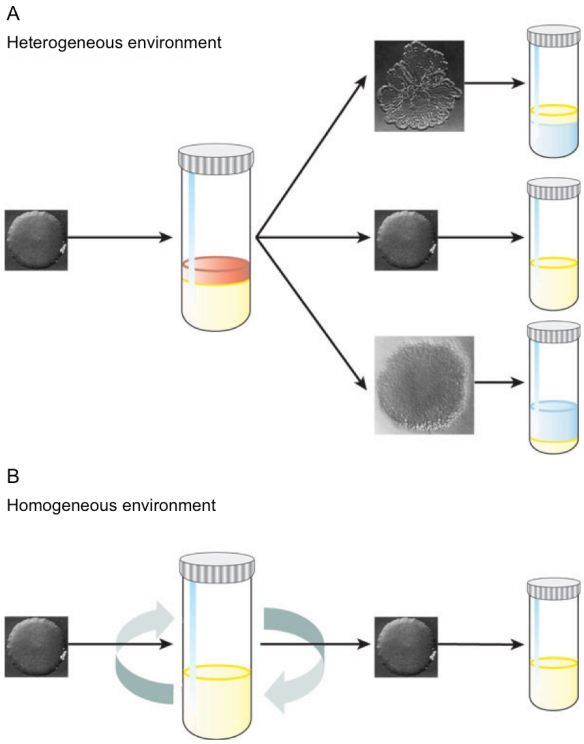


Figure 4|**Adaptive radiation in microcosms.** Experimental evolution of a single clone of *Pseudomonas fluorescens* in A) a heterogeneous environment and B) a homogeneous environment performed by Rainey and Travisano [24]. A) In a heterogeneous environment a single clone gave rise to three subpopulations each occupying distinct niches within the spatially structured microcosms. B) In contrast, no diverse morphologies were observed in a homologous environment. Figure is adapted from Buckling *et al.* (2009) [23].

Another example of diversity generation was demonstrated by Singh and colleagues who showed that *P. aeruginosa* rapidly undergoes extensive genetic diversification during short-term growth in biofilm communities as a consequence of a heterogeneous environment with nutrient gradients [25, 26].

Hence, it appears that complex spatially environments drive the emergence of niche specialists [22]. However, little is known about the diversity during long-term evolution and in natural environments. An interesting question is if the diversity is stably maintained over extended periods and if geographical locations of the populations maintain this diversity. We have addressed this in paper 1 (Chapter 6) where it was demonstrated that the geographical locations indeed play a fundamental role in shaping the population diversity in complex and spatially structured environments and that this generated diversity is stably maintained during long-term evolution [27].

2.1.3|Rewiring of regulatory networks

It has long been of great interest to understand the molecular mechanisms behind the phenotype changes during adaptation of an organism to a new environment.

Studies on molecular evolution have proposed that it is the regulatory genes, rather than structural genes, that are important during the adaption [28]. The reasons might be that mutations in transcription regulators can change the binding specificity or create or delete interactions and thereby shaping

entirely new traits [29]. In this way mutations in transcriptional regulation can alter the function or activity of one central protein involved in a complex regulatory network. Therefore, regulatory mutations often have widespread pleiotropic effects that are likely to have a fitness cost, and compensatory mutation may arise to modulate the regulatory network to reduce some of these side effects (Figure 5) [30].

Furthermore, several studies have shown that adaptive mutations in global regulatory genes involve major epistatic interactions [12, 31, 32]. Damkiær and colleagues demonstrated that epistatic interactions of global regulators played a fundamental role in adaptation of the dominant *P. aeruginosa* DK2 lineage to the cystic fibrosis environment [32].

Because bacterial genes are regulated by multiple transcription factors working cooperatively or competitively, transcriptional rewiring caused by mutations may result in the establishment of novel combinational regulation [28]. Likewise, we showed that mutation in a global regulator modified the regulon of this protein and resulted in significantly distinct gene expression profiles and catabolism compared to both a wild type and a knockout mutant (see Chapter 6, paper 2 [33]).

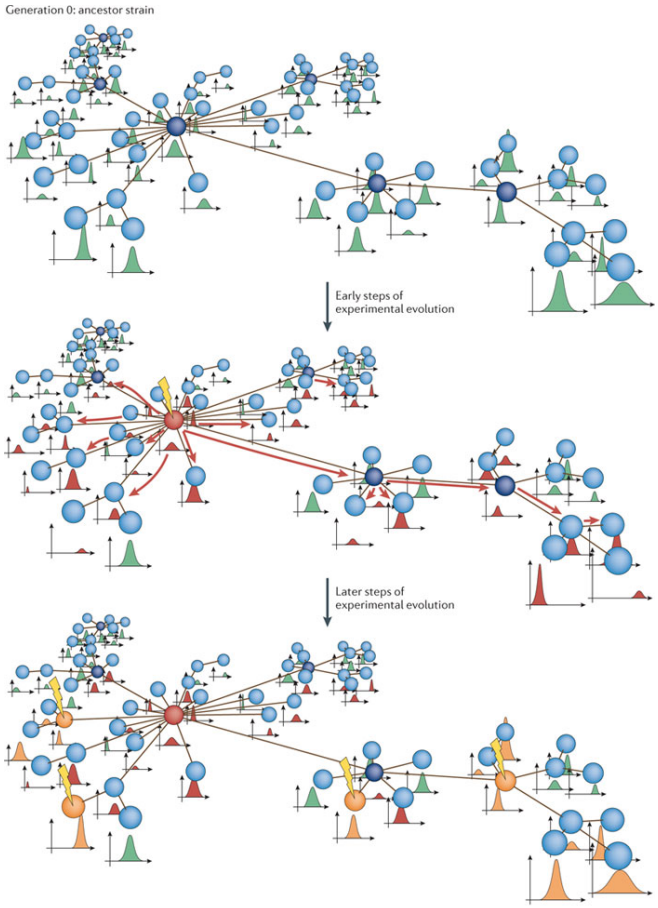


Figure 5|**Evolution of regulatory networks.** Example of a regulatory network (top). An early mutation in a central regulatory protein results in alteration of several other proteins where some are beneficial while others are not (middle). To account for this, later in evolution, compensatory mutation may arise (bottom). Figure is adapted from Hindre *et al*. (2012) [30].

Evolution through bacterial transcription rewiring is mainly focused on the interaction between transcription factors, RNA polymerase and/or target promoters, however, other factors affecting transcription exist including secondary messengers, regulatory RNA and small RNAs [28]. Due to the scope of this thesis, these factors will not be described here.

2.1.4|Hypermutators and their role in adaptation

During adaptation hypermutators have been frequently found to arise both under laboratory conditions [4] and in natural environments [34]. These hypermutators can have increased mutation rate up to 1000-fold caused by defects in DNA repair associated genes [35].

The exact role of mutators in bacterial evolution is unclear, however, models have shown that the prevalence of strong mutators genes, that increase the mutation rate, can accelerate adaptation in an evolving population [35, 36]. Consequently, mutators can result in a rapid acquisition of beneficial mutations that enable them to specialize in the environment, but may have reduced capacity to survive elsewhere and therefore reduce the transmissibility of the strain [37].

It has been described that mutators can spread and be fixed by hitchhiking with the beneficial mutations they generate [36, 38]. Likewise, in many studies the appearance of mutators have been linked to increased antibiotic resistance [34, 39]. Furthermore, complex and spatial habitats have been shown to promote evolution of mechanisms to enhance the rate of variation [34-36]. Under these conditions with strong selective challenges, bottlenecks are predicted to occur and thus giving mutators a significant role in bacterial evolution [34]. While in a constant environment, mutation rates have been predicted to remain low [38].

It has been shown that bacterial mutation rates change during adaptation so that a high mutation rate is beneficial at the beginning, but the benefit disappears when adaptation is reached [37]. Thus in highly adapted populations, mutators are expected to be rare, as they frequently will generate lethal or deleterious mutations. While in an evolving population, mutators accelerate adaptation by generating beneficial mutations [37].

2.2|Horizontal Gene Transfer in bacterial evolution

Sequential modifications of gene function by accumulation of point mutations have indeed contributed to the diversification and adaptation of microorganisms as described above. This kind of adaptation to very hostile and novel environments can be assumed to be a slow evolutionary process. Instead, horizontal gene transfer (HGT) also known as lateral transfer represents a very fast evolutionary process in which the recipient cell rapidly acquires new traits. The widespread of antibiotic resistance and virulence have long been associated with HGT and are a major challenges in the clinical settings [40].

There are two advantages with uptake of homologous DNA: the new DNA may repair damaged DNA on the host chromosome or the cell can acquire a new gene that confers a selective advantage. If the uptake is advantageous, the cell lineage may be spread and maintained in the population. Some HGT are not beneficial, but causes deleterious effects, resulting in eventual extinction of these cells. Finally, in a polymicrobial community HGT may contribute to the population diversity by generation of novel strains with mosaic genomes that blurs the boundaries between species [40] (Figure 6). During a chronic infectious process, HGT may therefore result in the continuous generation of a cloud of new strains with novel combination of genes and distinct phenotypes. This may lead to increased disease severity as HGT provides the bacterial population with means to impede the adaptive immune response and generate high diversity [41, 42]. For this matter HGT has an important role in bacterial evolution and population dynamics and it is therefore of great importance to detect and study HGT for better understanding of evolutionary processes.

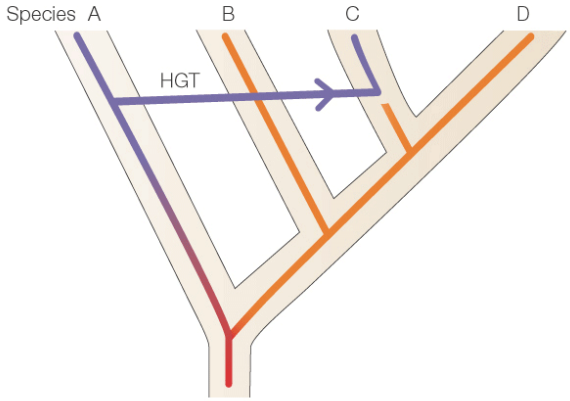


Figure 6|**Illustration of evolution of four species.** The main lines show the vertical evolution of the species. During evolution genes in each of the species evolve. At some point a set of genes is transferred from species A to C (horizontal line) generating a novel strain with mosaic genome that may blur the boundaries between species. Figure is modified from Gogarten and Townsend (2005) [43].

How can it be determined whether a new trait or genomic variation is a result of HGT? Outside of experimental settings, HGT are rarely detected. The most powerful tool to detect HGT is whole genome sequencing, since HGT creates an unusually high similarity between the donor and the recipient cell [44]. Several methods for analysis of genome sequences to identify cases of HGT have been used. Beside from searching for uneven distribution patterns and comparing of phylogenetic trees, unusual nucleotide composition (e.g. codon usage, GC content) can give information of the donor origin, as sequences retain the characteristics of the donor genome [45]. The availability of the increasing number of whole genomic sequences provides the opportunity to detect and determine the frequencies of lateral transfer in natural populations during adaptation. Likewise, we detected an *in vivo* HGT between two coexisting clones of *P. aeruginosa* sampled from a cystic fibrosis patient by identifying non-randomly distributed SNPs in the recipient clone and mapping them to the a possible donor clone (see paper 3 Chapter 6 [46]).

HGT can be considered a multistep process where several factors are necessary for the transfer to occur. First, there needs to be a supply of donor DNA. Second the DNA must be incorporated into the genome of the recipient cell. Third, the incorporated DNA must have reproductive advantage that benefits the recipient cell in order to persist. Delivery and incorporation of the DNA typically takes place through three mechanisms: transformation, transduction, or conjugation. Transformation is the uptake an incorporation of naked DNA. Conjugation is a cell contact-dependent DNA transfer mechanism. The transfer can be mediated by self-transmissible conjugative plasmid, mobilization, cointegration, or transposons. Transduction is mediated by bacteriophages that act as vector for injection of DNA into a recipient cell.

For each of the described steps limitations and barriers exist, including supply of DNA, establishment of cell-cell contact, cell competence, deleterious effects, etc.

With these limitations in mind, it has been demonstrated that all necessary elements for successful HGT exists under chronic infectious conditions [47]. Chronic bacterial infections are often associated with biofilms, where multiple species and phenotypes may coexist [41, 48] and form close interactions [49]. The surrounding protective matrix of the biofilm contains exopolysaccharides, proteins, lipids, and a high level of extracellular DNA (eDNA) from lysed or dead cells [50]. Moreover, transformation has been demonstrated to be 10^4-fold higher in biofilms than in planktonic forms [51]. These foundations make biofilm and chronic infectious conditions perfect settings for HGT [42].

Chapter 3

Evolution in natural environments

Most knowledge we have today on bacterial adaptation and evolution to novel environments is based on experimental evolution studies performed under defined laboratory conditions. These studies have provided great insight into the adaptive evolution of bacterial populations. What is characteristic for these evolutionary laboratory experiments are the lack of environmental complexity and spatial structure. Even though specific polymorphisms occasionally occur, there is a general lack of population diversity in the adapting cultures. Therefore, it is uncertain how relevant these findings are for the evolutionary dynamics in natural environments where limited information is available. Mimicking a natural environment is very difficult because of multiple selective pressures that fluctuate in both time and space shaping the evolution of an adapting strain.

For this reason, chronic infections with *P. aeruginosa* in cystic fibrosis patients give a unique opportunity to study the adaptation and evolution of a pathogen to the heterogeneous and fluctuating environment in the airway of patients with cystic fibrosis patients. Such studies may provide important information for future therapeutic strategies.

3.1|Cystic Fibrosis

Cystic fibrosis (CF) is the most common autosomal recessive disorder affecting Caucasians, with an incidence of approximately one in 2,500 individuls [52]. It is caused by mutations in the CF transmembrane regulator (CFTR) gene that encodes a chloride channel [53, 54]. The CFTR protein belongs to the ABC transporter (ATP-binding-cassette transporter) super family and functions as the primary chloride channel of epithelial cells located in the respiratory, gastrointestinal, hepatobiliary, and reproductive system [55]. CFTR and the epithelial Na^+ channel (ENaC) are very important for maintaining the salt-water balance across the epithelial cell membranes.

More than 1800 mutations in the CFTR gene have been identified [56]. However, the most common mutation is ΔF508 that accounts for approximately 70% of CF chromosomes worldwide and 80% in Danish CF patients [54, 55, 57, 58]. The mutation is a 3-bp deletion, resulting in loss of a phenylalanine at position 508 located in the nucleotide-binding domain 1 [54, 57]. Due to this mutation, the protein formed after translation fails to mature properly and is subsequently tagged for degradation, resulting in the loss of CFTR [59].

CFTR deletion or dysfunctionality results in abnormal transport of chloride and sodium across the epithelial cell membranes, dehydrating the airway surface liquid layer. As a consequence, thick dehydrated mucus is produced. The hyperviscous mucus is difficult to eradicate by mucociliary

clearance and coughing [60]. Due to this, opportunistic pathogens are able to evade and colonize the CF airways causing persistent, chronic infection. The deficient eradication of inhaled microbes by mucociliary clearance or coughing causes a continuously recruitment of polymorphonuclear leukocytes (PMNs) and thereby chronic inflammation [55].

3.1.1|Compartments of the CF airway

The human respiratory system can be divided into upper and lower airways. The upper airways consist of the paranasal sinuses, which are a group of air-filled-spaces. The maxillary sinuses are located around the nasal cavity and the frontal sinuses are placed above the eyes. The ethmoidal sinuses are many small sinuses located between the eyes, and the sphenoid sinuses are placed behind the ethmoids [61] (Figure 7). These small cavities facilitates the trapping of thick mucus in CF patients and it can be difficult to get access to the sinuses due to abnormal anatomy of CF patients.

The lower airway can be divided into the conductive and respiratory zones (Figure 7). The conductive zone refers to the trachea, bronchi, and terminal bronchioles. Cilia, submucosal glands and goblet cells are present in the conductive zone where also the mucus is produced [62]. In the conductive zone, infecting microbes are found mainly inside sputum, where conditions are anaerobic.

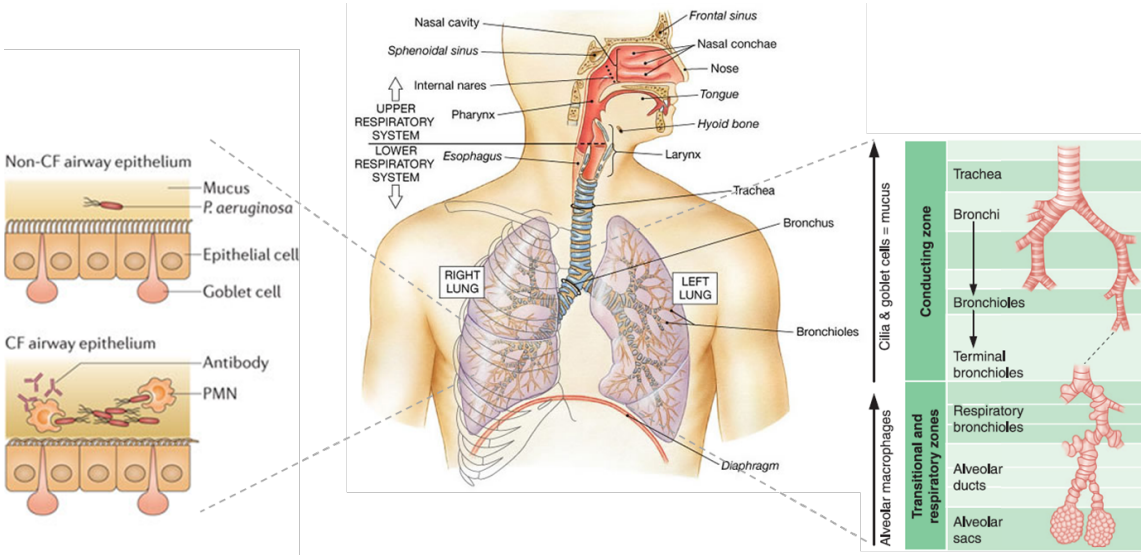


Figure 7|**The airway of CF patients.** The human respiratory system contains different anatomical compartments, including the paranasal sinuses, the conductive zone (tracheobronchial region), and the respiratory zone (alveolar region). Infecting microbes is removed in healthy subjects by the cilia of the epithelial cells, however, in CF patient thick dehydrated mucus is produces that enables bacteria such as *P. aeruginosa* to colonize. Continuous recruitment of PMNs results in increased inflammation and tissue damage. The airway mucus of the different compartment in the CF airway constitutes a reservoir for *P. aeruginosa* growth. Figure is modified from Høiby *et al.* (2006) and Folkesson *et al.* (2012) [62, 65].

The respiratory zone contains no cilia and submucosal glands, but includes the bronchioles and alveolar ducts and sacs, where the O_2 and CO_2 exchange takes place [62]. Therefore, bacteria in the respiratory zone are probably exposed to aerobic as well as microaerophilic growth conditions. The amount of PMNs is also higher here because of the blood vessels around alveoli [63, 64]. The presence of bacteria

is rare in the respiratory zones, but specific regions often becomes destroyed because of the inflammation surrounding infected areas [63].

Several studies have suggested that the initial infection starts in the upper airways [66-68]. After several years with infection in the sinuses, the same organisms are thought to migrate and colonize the lower airways. The transmission from upper to lower airway is probably through the larynx and trachea [68].

While, the small cavities of the sinuses facilitate trapping of the thick mucus, there may be a better clearance in the lung by breathing and coughing. It has been shown that the immune response is reduced in the sinuses compared to the lung [69, 70]. This means that the upper airway has less airflow, and less exposure to antibiotics and host immune cells than the lower airways. In this way, the sinuses may act as a reservoir for infecting bacteria and promote their adaptation [66, 69]. Consequently, early treatment of sinus infections may delay the development of significant chronic lung infection. Some hospitals require sinus surgery before lung transplantation to prevent infection in the lungs by bacteria colonizing the upper airway [68].

3.1.2|CF airway ecology

It is well documented that the major reasons for morbidity and mortality of CF patients is chronic lung infection caused by the colonizing microbes. The microbial community of CF patients has been found to change over time (Figure 8).

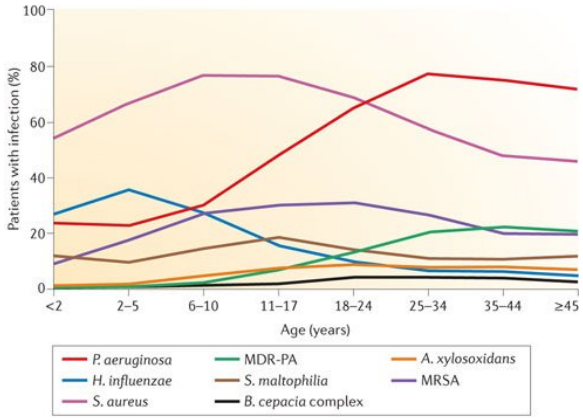


Figure 8|**Prevalence of common human respiratory pathogens in patients with cystic fibrosis as a function of age.** While *P. aeruginosa* is the most common pathogen in CF adults, *S. aureus* and *H. influenza* infections are common in early childhood. Figure adapted from Folkesson *et al.* (2012) [65].

As shown in Figure 8, *Staphylococcus aureus* and *Haemophilus influenza* infections are common in early childhood. These organisms damage the epithelial surfaces leading to increased attachment by *P. aeruginosa*, which is the major pathogen observed in adult CF patients [55].

Most CF patients are initially infected with organisms found in their immediate environment, but patient-to-patient transmission have been detected [71, 72]. Contact to other CF patients therefore constitutes cross-infection risk.

As described above, each spatial site of the CF habitat is a very complex and dynamic ecosystem, where many factors contribute to the stressful environment. Microorganisms need to adapt to these multiple challenges, including antibiotic exposure, the immune system, nutrient availability, microbial interactions and heterogenous oxygen levels that fluctuate both in time and space in order to survive [73]. Some of these will be described here.

Antibiotic treatment

One of the major challenges for the bacteria to adapt to the CF airways is the fluctuating concentration of antibiotics that they are exposed to over time. CF patient are often treated early and aggressively with antibiotics to eradicate or prevent chronic infection [74]. Inhaled antibiotics can result in high concentrations of the drug in the airway mucus, but low in the respiratory zone [62, 73]. The situation is reversed if the antibiotic is given intravenous [62, 73]. In this way, antibiotics can change the community structure and stability, but also promote the adaptation of strains with mutations conferring antibiotic resistance [73, 75, 76].

The immune system

The immune response constitutes another significant selective force that shapes the adaptation of infecting microorganisms. The continuously recruitment of PMNs and neutrophils by the invading microorganisms results in inflammation and tissue damage [55]. Furthermore, different eukaryotic-specific receptors located on the surface of epithelium cells can recognize bacterial epitopes, including lipopolysaccharide (LPS) and flagellin, leading to killing of the bacteria [77, 78]. Moreover, phagocytosis of bacteria produces reactive oxygen species (ROS) and microorganisms are therefore also encountered with high oxidative stress [62].

Nutrient availability

The physical and chemical properties of the CF environment are altered in comparison with healthy subjects. The availability of nutrients therefore constitutes another major selective force. The CF sputum is a complex substrate that provides the infecting bacteria with carbon and energy to support growth during colonization [79]. For example, the sputum of CF patients is rich in different amino acids [80] and it has been shown that the preferred carbon sources of clinical *P. aeruginosa* isolates sampled from CF patients were proline, alanine, arginine, lactate, glutamate, and aspartate [81], demonstrating that the nutrient environment impacts adaptation of infecting microbes.

Microbial interactions

Most CF infections are characterized by colonization of more than one species (Figure 8) [55]. Interactions between the different organisms of the microbial community also impose a major challenge for the adapting microbe. However, the effect on the individual behaviour and the population dynamic is very poorly understood. Because of the unavoidable competition for nutrients, antagonistic interactions

are common within a community [82]. Two examples of such antagonistic interaction are the production of bacteriocin (see section 3.2.2) and the breakdown of cooperative behaviours due to the occurrence of cheaters [82, 83]. A cheater refers to a non-cooperative strain that does not invest in the public good, but only utilize them. A classic example is the occurrence of siderophore-deficient mutants in a *P. aeruginosa* population. These mutants do not produce siderophores but are still able to use them to acquire iron from the environment [84].
Synergistic interactions have also been documented [85]. For example it has been demonstrated that *P. aeruginosa* show increased virulence when grown together with gram-positive bacteria [86]. Synergistic interaction can therefore result in increased disease severity [82].

Other factors such as the heterogeneous distribution of oxygen also play important roles in shaping the evolution of adapting bacteria to the CF airway. To survive in this stressful environment, microorganisms therefore have to overcome these challenges by phenotypic and genomic adaptation.

3.1.3|Infection stages

The morbidity and mortality of CF patients is a consequence of respiratory failure and chronic infection caused by the colonizing microorganisms [55]. Two criteria are used to distinguish between the different phases of infection: the Copenhagen criteria and the Leeds criteria.
Based on lower airway samples taken 10-12 times per year, the Copenhagen criteria grades pulmonary infection into 1) never been colonized, 2) intermittently colonized and 3) chronically infected. Most hospitals do not see the patients as regularly as in Copenhagen, and therefore uses the Leeds criteria where patients are seen every third month. The Leeds criteria grade the infection into 1) never been infected, meaning that there has never been growth of any CF related Gram-negative bacteria. 2) Non-infected where no growth of any CF related Gram-negative bacteria is observed during a year. 3) Intermittent colonization where growth is observed in >0% and ≤50% of the samples and 4) chronic infection where growth is observed in >50% of a patient's monthly lower-airway samples.
The chronic stage is characterized by constantly high serum levels of immunoglobulin G (IgG) antibodies and numerous PMNs in the lower airways [87].
Early, aggressive antibiotic treatment is important to prevent or postpone chronic infection. Recent studies have demonstrated that sinus surgery can postpone chronic infection [88], confirming the importance of the sinuses as bacterial reservoirs and their role in establishment of chronic pulmonary infections.

3.2|Pseudomonas aeruginosa

P. aeruginosa is a gram-negative, rod-shaped bacterium that primarily lives in water, soil, and vegetation. It is an opportunistic pathogen capable of infecting a variety of different organisms including

amoeba [89], plants [90], fruit flies [91], nematodes [92] and animals [93, 94]. *P. aeruginosa* infection in humans is generally restricted to immunocompromised patients such as patients suffering from burns, cancer, HIV and cystic fibrosis [95, 96].

P. aeruginosa is a motile bacterium that is able to swim, swarm and twitch using a flagellum and/or pili. Its optimum growth temperature is 37°C but is able to grow at 42°C. The genome is 5.2-7.0 Mbp with 5,570 open reading frames making it one of the largest genomes among bacteria [97]. *P. aeruginosa* encodes a large number of regulatory proteins that reflect the broad adaptability and metabolic versatility of this organism.

The minimal requirements of *P. aeruginosa* together with its large genome size and complex regulatory networks, makes it capable of growing in a diverse set of ecological habitats. Furthermore, *P. aeruginosa* has the ability to quickly switch lifestyle upon changing conditions. These abilities may have important consequences for humans, providing the possibility of contact and cross-infections [98].

The pathogenesis of *P. aeruginosa* rely on the production and interaction of multiple virulence factors such as cell-associated factors (pili, flagella and LPS), secreted virulence factors (proteases, siderophores, pyocyanin, rhamnolipids, hydrogen cyanide, *etc.*), Type III secretion system, biofilm formation, alginate production and intrinsic resistance to a broad spectrum of antimicrobial agents. Due to the relevance of this thesis, only some of these virulence factors will be addressed here.

3.2.1|Cell-to-cell communication in pathogenesis of *P. aeruginosa*

The success of *P. aeruginosa* to cause infections can be described to the ability of *P. aeruginosa* to produce a broad variety of different virulence factors required for acute infections (Figure 9). The majority of these extracellular factors are not constitutive expressed but regulated in a cell density-dependent manner termed quorum sensing (QS). QS is mediated by small diffusible signal molecules that interact and activate their transcriptional regulators when reaching a specific concentration correlated with specific cell densities of the population [99].

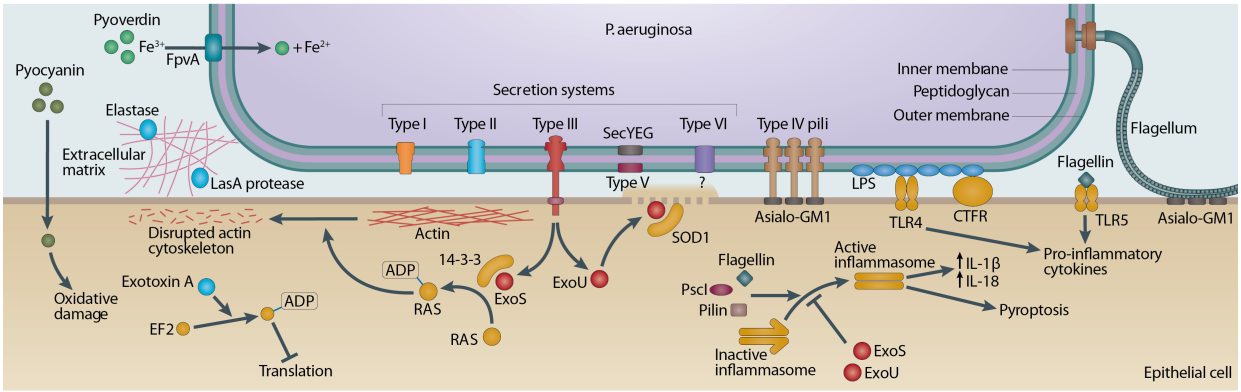


Figure 9|**Schematic overview of virulence factors of *P. aeruginosa*.** *P. aeruginosa* express multiple virulence factors important for acute infection including cell-associated factors, secreted virulence factors, and the Type III secretion system. The figure is adapted from Hauser and Ozer (2011) [100].

The *P. aeruginosa* QS system is mediated by two chemically distinct classes of signal molecules, the *N*-acyl-L-homoserine lactones (AHLs) [101] and the 4-quinolones [102, 103]. There are two AHL-dependent QS systems, namely the *las* system containing the transcriptional activator LasR and the AHL synthase LasI which directs the synthesis of *N*-3-oxo-dodecanoyl-homoserine lactone (oxo-C_{12}-HSL) [104], and the *rhl* system consisting of RhlR and RhlI which directs the synthesis of *N*-butanoyl-L-homoserine lactone (C_4-HSL) [105].

Beside AHL signaling, *P. aeruginosa* produces another signal belonging to the 4-quinolones called 3,4-dihydroxy-2-heptylquinoline referred to as PQS. PQS function as a link between *las* and *rhl* systems as demonstrated in Figure 10 and adds a level of complexity to the QS network. The QS systems are hieratically organized so that *las* transcriptionally regulates the *rhl* system [101, 106], and both *las* and *rhl* regulates PQS [103].

The QS system regulates multiple virulence factors through the regulation of these major systems, including production of elastase, exotoxin A, rhamnolipids, pyocyanin, hydrogen cyanide, siderophores, and biofilm formation [106, 107].

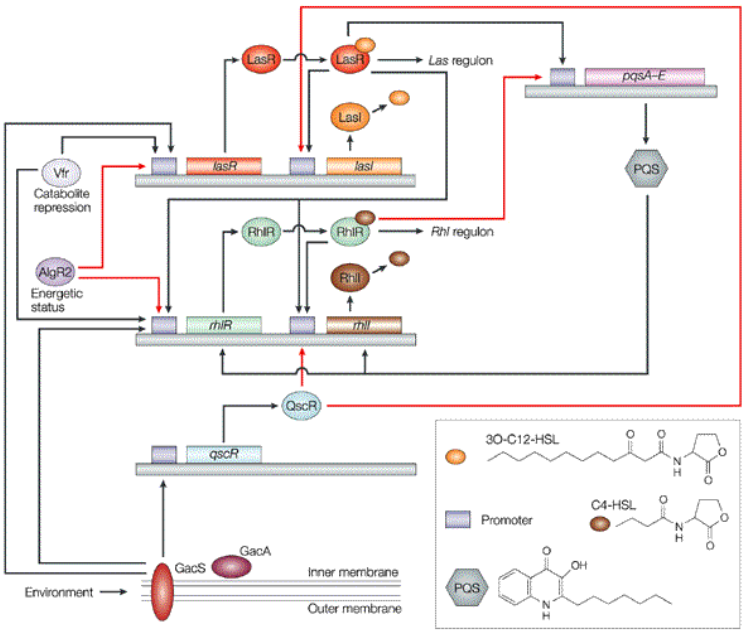


Figure 10|**A simplified overview of the quorum sensing regulation in *P. aeruginosa*.** Black arrows indicate positive regulation and red arrows indicate negative regulation. The *las*, *rhl*, and PQS QS system show a complex regulatory network. Adapted from Lazdunski *et al.* (2004) [108].

To further add a level of complexity, various regulators have been found to modulate the QS circuitry [109]. One of these is the virulence factor regulator (Vfr), an *E. coli* CRP homolog, which has been shown to control expression of more than 200 genes [110, 111]. Vfr directly activates *lasR* transcription by binding to Vfr-binding sequences in the *lasR* promoter region [112]. Moreover, Vfr has been demonstrated to directly regulate *rhlR* transcription by binding to several Vfr-binding sites present in

the *rhlR* promoter region, one of which negatively regulates *rhlR* [113]. In this way, Vfr regulates the production of multiple virulence factors important for acute infection mainly through the regulation of the early steps of the QS signalling cascade. Furthermore, Vfr has been indicated to negatively regulate the PQS system [114]. This complexity and the involvement of multiple regulatory proteins enables *P. aeruginosa* to modify the QS circuitry upon environmental changes and allowing adaptation to different environmental conditions.

3.2.2|Lipopolysaccharides and pyocins

P. aeruginosa produces an arsenal of different virulence factors that contribute to survival and protection against the immune systems, competing bacteria and aggressive antibiotic therapy. One of these major virulence factors is the cell wall-associated lipopolysaccharide (LPS). LPS is an essential component of the outer membrane and as the first line of defences it play fundamental roles for direct interaction with host cell receptors and antibiotics [115, 116].

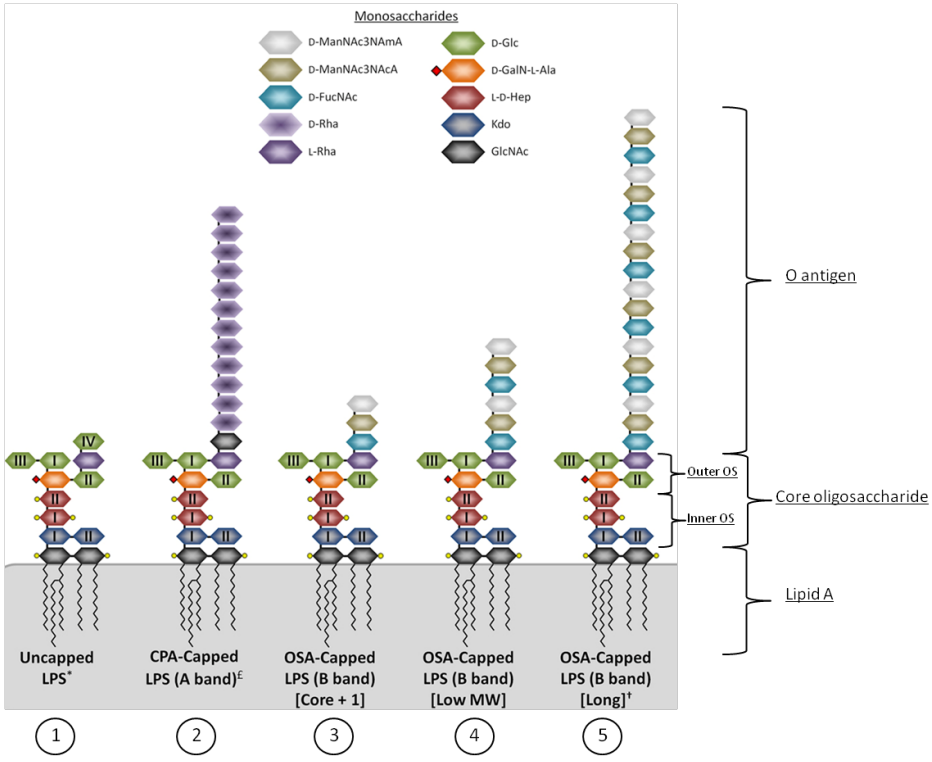


Figure 11|**Structures of LPS in *P. aeruginosa*.** LPS consist of three regions: lipid A, which anchors the LPS to the membrane, an O-antigen (O-Ag), and a core oligosaccharide (OS), which links the O-Ag to the lipid A. The OS can be either capped (O-Ag is attached, structure 1) or uncapped (no O-Ag is attached, structure 2-5), while the O-Ag can be expressed as two forms, namely A-band (structure 2) or B-bands (structure 3-5). The figure is modified from Lam *et al.* (2011) [115].

LPS of *P. aeruginosa* is composed of three distinct regions (Figure 11): lipid A, core oligosaccharide (OS), and the long-chain O antigen (O-Ag). The lipid A domain anchors LPS in the bacterial outer membrane by several fatty acid chains attached to a disaccharide backbone. The core OS is the linker between O-Ag and lipid A and it can be divided into an inner and an outer core. The inner core is highly conserved

comparing to other gram-negative bacteria and the composition of the sugars are identical among *P. aeruginosa* strains [115].

The outer core is divided into two structurally distinct glycoforms called 'capped' and 'uncapped'. The capped form is covalently bound to the O-Ag, while the uncapped version prevents attachment of O-Ag to the core OS [115, 117] (Figure 11).

P. aeruginosa is able to co-express two forms of O-Ag; the Common Polysaccharide Antigen (A-band also known as CPA) and O-Specific Antigen (B-band also known as OSA) [118, 119]. The B-band polysaccharides are the basis for the International Antigenic Typing Scheme (IATS) where *P. aeruginosa* is classified into 20 major serotypes (O1-O20) based on the variation in B-band structure and chain length.

While A-band LPS appears not to play a role in protecting the organism against serum-mediated killing, the B-band polysaccharides are highly immunogenic and have been found to confer serum resistance [120]. The LPS of *P. aeruginosa* have been shown to undergo phenotypic modification during chronic infection, so that A-band LPS becomes the major LPS antigen replacing the B-band [121-123]. It has been implicated that prolonged antibiotic treatment has caused this conversion [124]. Resistance to aminoglycosides is partly due to alterations in the LPS structure, resulting in decreased antibiotic passage across the outer membrane [125, 126].

Besides interaction with antibiotics and immune host cells, LPS has been described to function as receptor for R-pyocins [127-129]. In particular the sugar residues of the outer core has been suggested to be involved in interaction with R-pyocins [127, 129] (Figure 12).

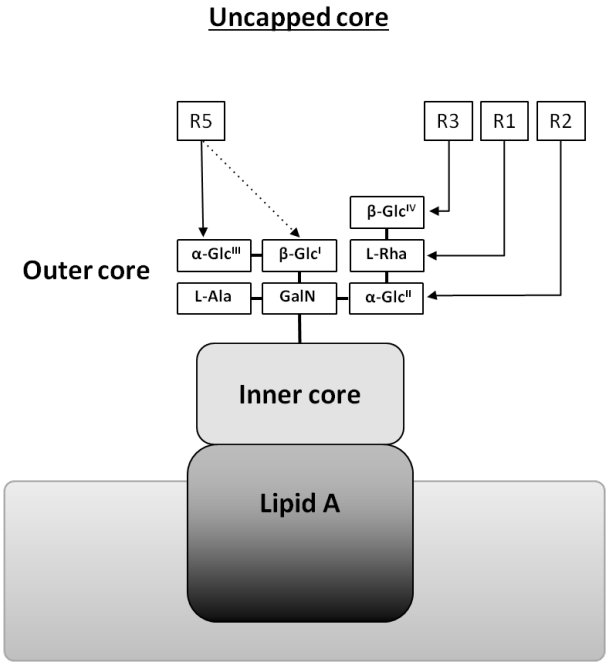


Figure 12|**Overview of the sugar residues involved in R-pyocin recognition.** Arrows indicate the sugar residues involved in R-pyocin interaction. R5 pyocin seems to be able recognize two sugar residues, but favour α-Glc^{III} over β-Glc^{I} (dashed line) [127].

Pyocins are bacteriocins and *P. aeruginosa* produces three types of pyocins: the soluble S-pyocin, the flexible F-pyocin and the non-flexible and contractive R-pyocin that resemble the tail of bacteriophages [130]. R-pyocins cause depolarization of the cytoplasmic membrane leading to cell lysis [119]. The killing efficiency is very high for R-pyocins as one pyocin molecule kills one bacterial cell. In comparison, 100-200 molecules of F-pyocins are required to kill one cell [130].

Five different types of R-pyocins have been identified based on their killing activities and difference in their amino acid sequence of the tail fiber protein Prf15 (PA0621) [127, 131].

Pyocins are able to kill members of the same closely related species and in this way ensure the predominance of a given strain in a niche [130, 132].

In summary, LPS play a fundamental role both as a protective shield and as receptors for multiple foreign molecules including antibiotics, immune system, competing microbes, pyocins, *etc.* Thus, LPS in clinical isolates is often modified [122, 133], conferring increased resistance and survival.

3.3|*P. aeruginosa* lifestyle and adaptation in CF airways

The CF airway infection offers a unique opportunity to study the lifestyle and adaptation of *P. aeruginosa* to a heterogeneous natural environment. Various factors contribute to the complexity of the ecosystem as discussed in section 3.1.2 that drives the evolution of *P. aeruginosa* to live in the CF ecosystem.

Adaptation of *P. aeruginosa* often follows a characteristic pattern with the initial colonization of *P. aeruginosa* strains. These strains often resemble environmental strains displaying wild-type phenotypes i.e. non-mucoid, fast growth, and susceptibility to antibiotics [134]. However, in the complex and highly dynamic CF environment where multiple selective pressures act on the colonizing *P. aeruginosa*, it continuously need to adapt in order to survive. Therefore, *P. aeruginosa* undergoes significant genomic and phenotypic changes to adjust to the novel conditions of the CF airways.

Parallel evolution towards some common traits referred to as 'the chronic phenotypes' have been shown in different clinical settings [39, 75, 76]. The genomic route to reach these end-point phenotypes can be very distinct [135].

3.3.1|Phenotypic adaptation

P. aeruginosa is able to rapidly switch lifestyle in response to changing environmental conditions. The phenotype changes significantly as a consequence of the selective pressures present in the environment during long-term evolution of *P. aeruginosa* in the CF airways. For example, estimations of *P. aeruginosa* growth rates in the CF airways show that *P. aeruginosa* populations have significantly decreased *in vitro* growth rate but are still actively growing *in vivo* [1, 73]. Furthermore, chronic infections with *P. aeruginosa* are often characterized by a biofilm lifestyle [136].

The biofilm serves as a protective physical barrier against the host immune response as well as antibiotic therapy and therefore display increased resistance to clearance and antibiotic resistance [137]. Biofilms can be considered as a dynamic living organism in which multiple species may coexist in a spatially structured environment, where they can mutate and evolve to improve interaction with other species [42, 138]. Furthermore, it has been suggested that self-generated diversity in biofilms may function as a form of biological insurance that protects the community unstable environments [25].

The most common phenotypes associated with long-term evolution in CF airways are antibiotic resistance [139], mucoidy [140], changed LPS [122], occurrence of 'small colony variants' [141], increased mutation rate (mutators) [34], QS loss [142], auxotrophy [80], motility loss and reduced production of acute virulence factors [75].

These phenotypes are remarkably different from the original invading strain and are observed repeatedly in different patients and in distinct clinical settings. The common reoccurrence of these traits suggests that there is a conserved pattern of evolution that *P. aeruginosa* follows in order to persist in the CF airways [75, 143]. Although there seem to be a parallel evolution toward chronic phenotypes, it does not exclude phenotypic diversity among clinical isolates. In fact, high phenotypic diversity has been observed even between isolates from the same sputum sample, where multiple phenotypes of the above mentioned variants could be isolated [69, 144-146]. It is not known whether this adaptive radiation is a consequence of niche specialization or if the different subpopulations are homogenously distributed in the CF airway (Figure 13). However, in paper 1 of this thesis, we addressed this important unanswered question and demonstrated that sub-lineages were located at distinct geographical niches within the infected airways [27].

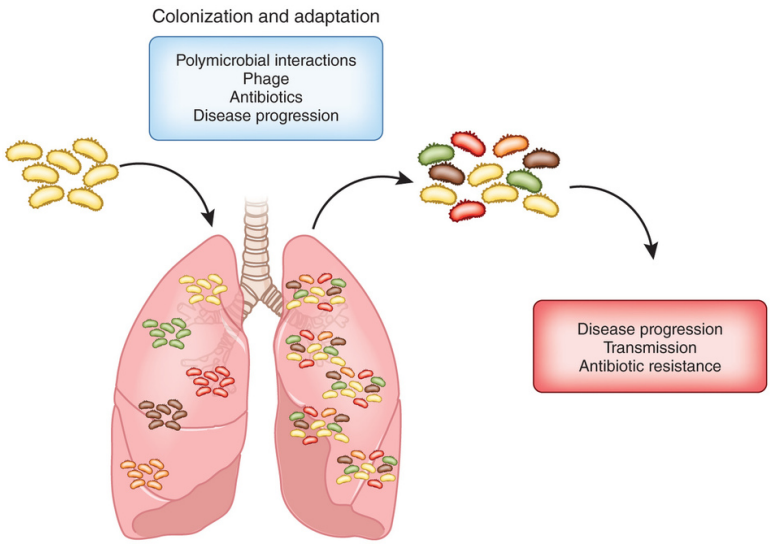


Figure 13|**Population heterogeneity of *P. Aeruginosa* within CF airways.** Spatial and temporal heterogeneity may drive and maintain diversification of *P. Aeruginosa* population. It is not known whether the population is geographically isolated (left lung image) or homogeneous distributed throughout the airways (right side of the lung image). Figure is adapted from Bernier *et al.* (2014) [147].

3.3.2|Genomic adaptation of *P. aeruginosa*

Although the chronic phenotypes are commonly observed, the underlying genomic changes may be very different. Little is known about the evolutionary routes used by *P. aeruginosa* to reach the phenotypic fitness peak in the fitness landscape. Whole genome sequencing has made it possible to detect and identify adaptive mutations acquired during chronic CF infection. In doing so a number of repeatedly observed mutations have been found. Mutations are in particular identified in the regulatory genes *mucA* and *lasR* [75, 76]. Inactivation of MucA results in overproduction of alginate leading to conversion to the mucoid phenotype [148]. This slimy biofilm-like phenotype is often observed in chronic infections and has several *in vitro* effects, including phagocytosis inhibition [149, 150], protection from reactive oxygen intermediates [151], and is further suggested to improve resistance to the innate clearance mechanisms in the lung [152].

Mutations in LasR influence the production of several virulence factors and confer growth advantages on certain amino acids [142]. Furthermore, genes important for antibiotic susceptibility, such as the efflux pumps, are also commonly mutated [75, 153]. Epistatic interactions between these mutations play important roles in the adaptation of *P. aeruginosa* [32].

Most studies identifying these adaptive mutations are based on few isolates from different time points representing the whole *P. aeruginosa* population [75, 154] and therefore conclusions about the population structure should be considered with caution. The extent of diversity and the ecological dynamics within chronic infection are not fully understood. Two models have been suggested: the 'dominant-lineage' model, which suggests that little diversity is observed in the population, and a 'diverse-community' model, where several subpopulations may coexist and compete [155, 156] (Figure 14).

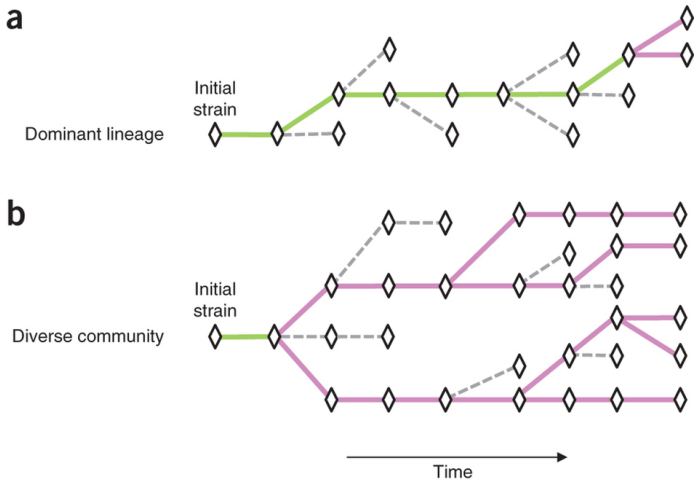


Figure 14|**Models of within-patient evolution.** A) In the 'Dominant-lineage' model beneficial mutations are fixed leading to one dominant lineage (green lines), while less fit ancestors are outcompeted (dashed lines). Polymorphic mutations will be rare in this model (magenta lines) [156]. B) In the 'Diverse-community' model co-existing subpopulations arise from the ancestor that may acquire distinct adaptive mutations leading to continuous diversification. Each subpopulation may climb local adaptive peaks resulting in niche specialisation. Figure is adapted from Lieberman *et al.* (2014) [156].

A heterogeneous environment with distinct niches have been shown to promote diversity [24]. Indeed the CF airway constitute the complexity that generates diversity and ecologists have long recognized that population diversity can enhance survival and drive persistence in chronic infections [41]. This is also known as the 'insurance hypothesis', which postulates that the presence of diverse subpopulations increases the range of conditions in which some will survive [157]. Therefore, within-patient population diversity has major implications for both for drug treatment and resistance [158], but also for understanding evolutionary processes and natural genetic microvariation [159].

Chapter 4

Present investigations

Studying natural evolution is very complex. Various challenges exist including sampling of populations in an organized manner, the environment may be problematic to define or the population of interest is too small. Few systems are therefore suited for such studies. The airway infections of *P. aeruginosa* in patients suffering from CF constitute an ideal model. The work presented in this thesis has only been possible to perform due to an unique collection of frozen 'fossils' of *P. aeruginosa* isolates derived from CF patients since 1973 at the Copenhagen CF Clinic [160]. This more than 40 year old strain collection provide an unique opportunity to design longitudinal investigations of evolutionary trajectories for different bacterial clones in distinct patients covering more than 150,000 bacterial generations [1].

4.1|Background

In a molecular epidemiological study it was shown that two dominant clones were repeatedly isolated from the CF patients in Copenhagen, namely DK1 (formerly known as genotype r) and DK2 (formerly known as genotype b) [161]. These two clones have successfully colonized the airways of multiple CF patients since 1973 and were found to be transmissible between patients [161] (Figure 15). Furthermore, in some patients the two clones appear to coexist, while in others, they seem to compete, because coexistence often resulted in niche exclusion and eradication of one of the clones [161]. The specific mechanisms involved in this competition is not known, however, the nutritional conditions, antibiotic treatment, the immune system, and other stress factors might have contributed to this competitive effect.

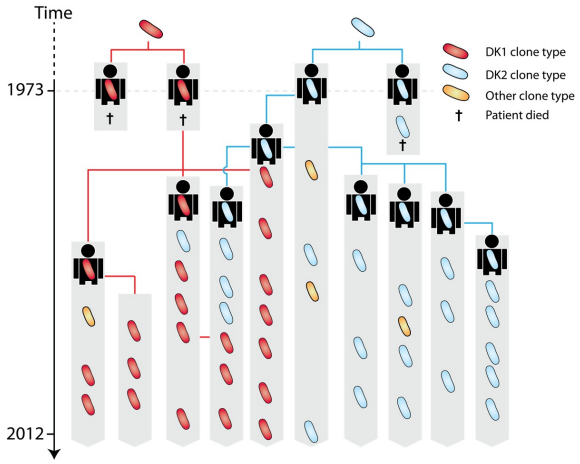


Figure 15|**Simplified overview of the epidemiology of the DK1 and DK2 clones of *P. aeruginosa* in Danish CF patients attending the Copenhagen CF Centre.** The DK1 and DK2 clones was first isolated from CF patients in the early 70'ties and have since then been transmitted and isolated from more than 40 patient each. Figure is adapted from [162].

The evolutionary trajectories and dynamic of the DK2 lineage has recently been investigated [76, 163]. Surprisingly, it was found that this dominant lineage evolved into a highly successful colonizer by acquiring few early pleiotropic mutations in regulatory genes that resulted in rewiring of global regulatory networks [32, 76]. The genomic signature indicated negative selection and it was observed that the adaptation leading to persistence of the DK2 clone did not lead to increased population diversity. One dominant cell line has been able to colonize all major niches of the CF airways in several different hosts [76]. These findings were indeed striking as they contradict previous observations of positive selection from long-term *in vitro* evolution experiments [164] and generation of diversity reflected by the presence of multiple niches [22]. A relevant and important question asked in this thesis is whether other clones of *P. aeruginosa* share this evolutionary path, or if alternative routes can be used.

In many ways, the CF airway model resembles the Lenksi evolutionary experiment. However, the human airway constitute a far more complex natural environment with a spatially structure that fluctuates in both time and space compared to Lenski's experimental setup [165].

4.2|Thesis objectives and questions

With the important and surprising findings of the evolutionary dynamics of DK2, we analysed the evolutionary dynamics of the DK1 clone.

The major objective of this thesis has therefore been to document how microbial populations use different evolutionary strategies when adapting to novel environments over long time periods. Furthermore, we investigated the underlying key mutational events, and mapped the subsequent functional progression of adaptation resulting in highly successful and transmissible cell lines with potential to persist indefinitely.

A key motivation in this thesis was to document cases where generation of population diversity is evident and address a number of important questions in evolutionary biology, including: is spatial structure of the environment causing polymorphism? Is the diversity stably maintained over extended time periods? And is diversity maintained through interactions between variant subpopulations?

The longitudinal collection of isolates from CF patients makes a perfect foundation to study this. With focus on DK1 we aimed at addressing these and others important questions in evolutionary biology by the presented papers in Chapter 6.

4.2|Outline of studies

Paper 1: In the first paper, we examined the temporal and spatial variations in the *P. aeruginosa* population diversity in a chronically infected individual during a 32-year period to gain new insights into the bacterial diversification processes.

Genome sequencing and phenotypic assays, including transcriptome analysis (Affymetrix GeneChiP) and catabolic performance (BIOLOG Phenotype Microarrays) were used to characterize the evolutionary trajectories and demonstrated that multiple pathways for adaptation existed for the ancestral genotype. We observed that the original infecting clone rapidly diversified into three distinct sub-lineages that co-existed for decades each with their own functional and genomic signatures and rates of adaptation. Using unique sampling materials, we demonstrated that the subpopulations occupy separate geographical niches each with different selective pressures in the infected airways (Figure 16). Based on our phenotypic and genomic data, we suggest that sinuses functioned as a reservoir for subsequent colonization of other available niches.

These data indicated that there is an underlying complex population structure in *P. aeruginosa* infections not previously realized represented as the long-term and stable intraclonal diversity.

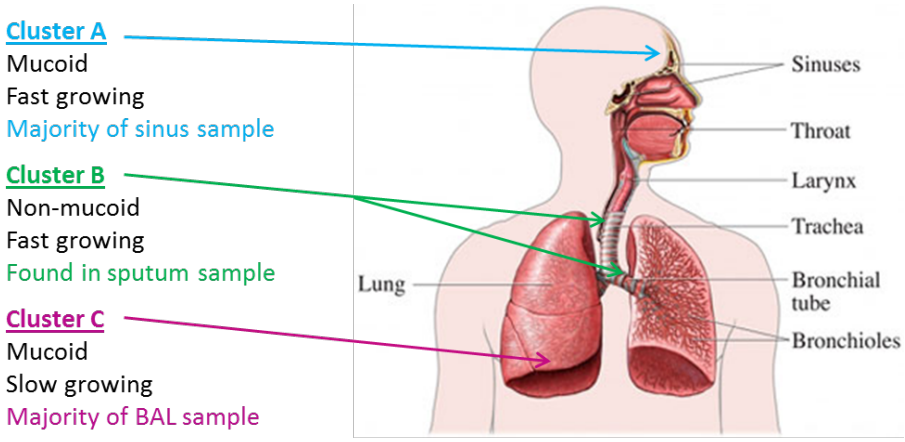


Figure 16|**Geographical location of three distinct subpopulations of *P. aeruginosa* in the CF airways.** The spatial structure were found to give rise to three different sub-lineages (A, B, and C) that were stably maintained during long-term evolution [27]. Isolates from cluster A was found to colonize the sinuses, while isolates from cluster C was sampled from a bronchoalveolar lavage (BAL). Isolates from cluster B was found in sputum together with isolates from cluster A and C.

Paper 2: In the second study [33], we investigated one of the adaptive mutations observed in the DK1 lineage. As mentioned, the DK2 lineage was found to early acquire important regulatory mutations that rewired the regulatory network and played a fundamental role in adaptation. These important regulatory mutations were not observed in the DK1 lineage (with the exception of *mucA*). Instead, we observed an early mutation in the global transcriptional regulator Vfr (*vfr*^{T163P}). The mutation was predicted to occur before 1973 as all isolates with the exception of one 1973 isolate had the mutation. For this reason and with previously results from DK2 in mind, this mutation was of great interest, and was hypothesized to have a central role in the adaptation of *P. aeruginosa* to the CF environment.

To clarify the role of *vfr*^{T163P} we transferred it to a PAO1 background and characterized regulatory consequences and pleiotropic effects caused by the mutation compared to a wild type and a knockout mutant. We discovered that the mutation occurred in the RNA polymerase interaction region and that it

resulted in modification of several phenotypes including transcriptional and catabolic profiles. The *vfr*^{T163P} phenotype resembled a chronic infection phenotype. This indicated that a single naturally occurring mutation caused completely alterations of multiple phenotypes reflecting the transition from an environmental strain to a host specific pathogen for CF patients.

Paper 3: The CF airways are often colonized by multi-species. Interactions between different clone types are one of the important selective pressures that shape bacterial evolution. The contribution of clone interaction to both population diversity and further evolution, but also patient morbidity, remains largely unelucidated.

In study 3, we identified an *in vivo* HGT via whole genome sequencing in a CF patient polyclonal infected with *P. aeruginosa* DK1 and DK2 clone. To gain new insight into the evolutionary roles of HGT during chronic polyclonal infections, we examined the population structure and dynamics of this CF patient. We elucidated that the two clones were colonizing the same niche in order for the HGT to occur and demonstrated that multiple subpopulations of both DK1 and DK2 co-existed in the spatially and complex CF airway. The HGT contributed to this diversity by generating novel strains with mosaic genomes that was found to persist in the CF airways.

Finally, we showed that the HGT resulted in altered intraclonal interactions by increasing the resistance to R5 pyocin giving the recipient cell a selective advantage in bacterial warfare.

Chapter 5

Conclusion and future perspective

The study of bacterial adaption to novel environments can provide great insight into the evolutionary processes and selective pressures driving the evolution. A major interest is to understand the mechanisms that promote and maintain genetic diversity within ecosystems.

In this thesis, we investigated the diversity generation of evolving microbial populations using the *P. aeruginosa* airway infection in CF patient as a model system. Although *P. aeruginosa* population diversity has been demonstrated by investigation of bacteria cultured from the same respiratory specimens, the understanding of how this diversity is generated and maintained is limited. Here we demonstrated two cases on how this diversity can be generated.

With focus on the dominant and highly successful DK1 clone, we presented a case in which DK1 evolved in the highly complex and heterogeneous CF environment of one patient for 32 years. We demonstrated that not only did the ancestral clone diversify into several subpopulations rapidly after colonization, but we were also able to show that these subpopulations, with distinct genomic and phenotypic features, were stably maintained during long-term infection. Therefore, sequencing a single isolate to represent the whole population would dramatically underestimate the intraclonal diversity and such results should be interpreted with caution. Moreover, it was clearly demonstrated that the subpopulations occupied distinct geographical locations within the CF airways.

These data demonstrate that stable diversification within *P. aeruginosa* population can be caused by spatial compartmentalization in the complex CF environment.

It was shown that before diversification, the DK1 clone acquired an early regulatory mutation in a global transcriptional regulator (Vfr). The Vfr is implicated in the complex network of regulating acute virulence factors mainly through regulation of *lasR*. Mutation in this regulator lead to alteration of multiple phenotypes reflecting the transition from an environmental strain to a chronic strain. This indicated that mutation in *vfr* was the first step of becoming a successful colonizer.

Another case of diversification was observed in a patient polyclonal infected with DK1 and DK2 for more than one decade. Via a genomic approach we identified an *in vivo* HGT between the two clones that significantly contributed to the generation of population diversity. The novel strain (termed DK1/2) was found to persist together with both DK1 and DK2 in the CF airways. Moreover, the transfer was shown to give the recipient cell an increased species-specific-antibiotic (pyocin) resistance.

In conclusion, adaptation of DK1 to the highly structured and heterogenous CF environments is facilitated by high intraclonal diversity and niche specialization. These results point toward an

underlying complex population structure not previous realized, and show that the spatially structured CF environment and composition, and interaction between species within the community play significantly roles in diversification of the population.

5.1|Different evolutionary paths of DK1 and DK2

The investigation of the DK1 clone revealed a distinct evolutionary path compared to that previously observed for the other dominating lineage, DK2. Although the two clones have evolved under similar selective pressures and antibiotic treatment and in some cases even in the same CF patients, it was demonstrated that the two clones evolved differently. Previous study documented limited diversification of the DK2 lineage despite the stressful and fluctuating CF environment [76].

In contrast, we observed high degree of diversification and niche specialization of the DK1 clone in the case of isolates from patient P30M0 (paper 1). To further examine the DK1 population dynamics and to exclude that the observed diversity in P30M0 was a special case, we have genome sequenced 12 additional isolates from multiple patients infected with the DK1 clone. These preliminary results confirm the high diversity within the DK1 population (Figure 17) and show a population with high genomic heterogeneity, occurrence of mutators, and deep branching resulting in phenotypic variation, co-existing subpopulations and niche specialization. We therefore suggest that the two clones evolved through distinct evolutionary routes; DK1 have evolved into a niche specialist where co-existing subpopulations each climbed a distinct adaptive peak in the fitness landscape (illustrated as the 'diverse-community' model in Figure 14b), while the DK2 lineage reached an adaptive peak in the fitness landscape that made it possible to colonize all major niches of the CF airways and in several different hosts as a generalist. This type of evolutionary path may represent the 'dominant-lineage' model illustrated in (Figure 14a).

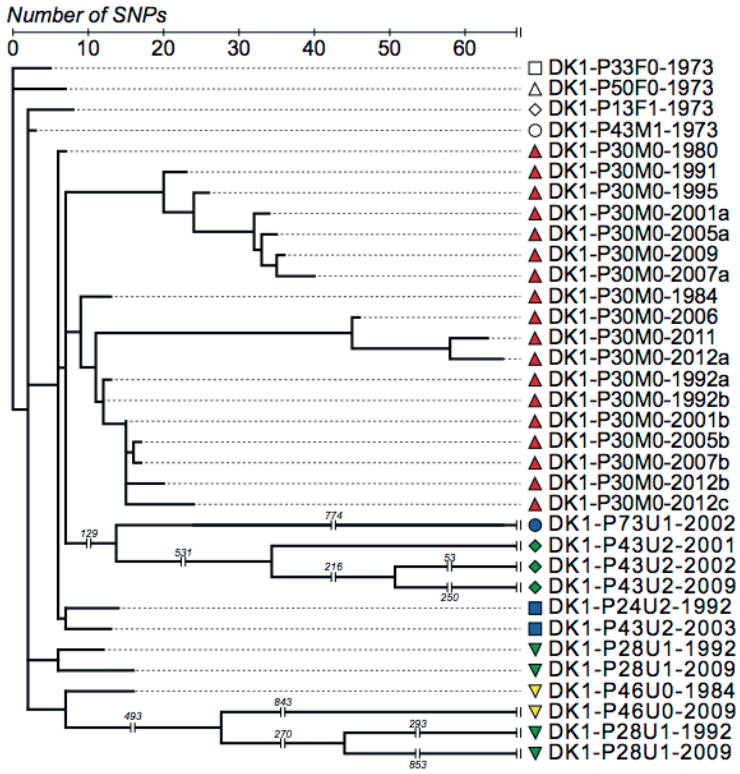


Figure **17|Evolutionary trajectories of *P. aeruginosa* DK1 clone.** Maximum-parsimonious reconstruction of the DK1 clones isolated from multiple patients. The phylogenetic tree is based on 6,362 SNPs identified from whole-genome sequencing and length of branches are proportional to the number of mutations. Broken branches indicate hypermutators and the above number represent the number of accumulated SNPs in the given branch. Mutation detection, analysis and construction of the phylogenetic tree were performed as previously described [163].

A limitation of the DK2 study is that only one contemporary isolate from each sputum samples is sequenced and compared and major conclusions on the whole population structure should therefore be considered with caution.

Other studies have demonstrated intraclonal population diversity of *P. aeruginosa* [69, 159] but it has also been reported for other pathogens including *Burkholderia cepacia* complex [156, 166] and *Helicobacter pylori* [167] when adapting to the CF environment. This indicates that diversification within patients during adaptation is the rule rather than the exception, and that DK2 might be a rare case. However, in a recent study by Marvig *et al* (2013), the possibility of intraclonal diversity in the DK2 was indicated [163] and we further show a case in a polyclonal infection where the population structure of DK2 is very complex [46]. As a next step, we therefore suggest to study the intraclonal diversity in depth by sequencing multiple contemporary and temporal sputum samples of both DK1 and DK2 to confirm the observed diversity (DK1) or the lack of the same (DK2).

Another intriguing question to ask is if one of the clones is more fit than the other, since we in several cases have observed coexistence of the two clones. Data indicate that they are colonizing the same niches [46, 161] and most often this coexistence results in eradication of one of the clones, indicating that they are competing in the CF airways [161]. Therefore, a head-to-head fitness experiment could be

very interesting. Unfortunately, we lack an *in vitro* model system that can mimic the complexity of the CF environment and the selective pressures acting on the population. The use of synthetic cystic fibrosis sputum medium [79] have shown parallel selection of several mutations previously identified from patient isolates [168] and may therefore be a good way to start.

5.2|Perspectives in relation to evolutionary microbiology

It is now feasible via affordable, high-throughput sequencing to uncover the underlying genetic mechanisms and changes during adaptation of evolving microbial populations in novel environments through the direct sequencing of clinical isolates. The Copenhagen strain collection provides the unique opportunity to study the longest natural microbial evolution 'experiment' ever reported. It has been found that the evolutionary dynamics of *P. aeruginosa* resemble evolutionary dynamics documented from *in vitro* experiment [76, 164, 169]. In this way, parallels can be drawn from laboratory studies and the real-world evolution trajectories.

Several parallels have been observed between the major pathogens of CF when adapting to the CF environment [170]. Both *P. aeruginosa* [76, 163] and *Burkholderia dolosa* [171] acquire mutations in an overlapping set of genes such as *gyrA* (drug resistance), global regulators, and genes involved in iron scavenging and outer membrane proteins [156, 163]. Besides this, intraclonal diversity and coexisting subpopulations have been observed for both pathogens [27, 156].

Studying how pathogens adapt to their human host can in this way provide unique information into the evolutionary trajectories of pathogens. These examples illustrate the importance of studying the molecular details of evolving populations in natural environments in addition to populations in laboratory systems.

It is evident that most infections result from colonization by more than one microbe. How these pathogens and their evolved mutants may interact within the community is unclear. However, most laboratory studies have only been focusing on single bacterial species grown in isolation. With the identification of HGT between two dominant clones of *P. aeruginosa* we have seen a glimpse of these complex interactions between the infecting microbial communities. Therefore, it would be highly interesting to study the effect and interactions in the community not only between *P. aeruginosa* strains but also between the other infecting strains. This will give a better understanding of microbial evolution and allow deciphering inter-species interactions that can help in the search for novel or improved treatments.

The unique collection of *P. aeruginosa* isolated from Danish CF patients, were sampled after the onset of chronic infection. With the findings that the sinuses may act as a bacterial reservoir where they can adapt and evolve long before the onset of chronic pulmonary infections [27, 66, 69], it leads to the question what happens during early and intermittent infections both in the lung but also in the sinuses?

The next step is to investigate the isolates from very young and intermittently colonized patients. This will lead to identification of mutations that are immediate important for colonization and transition from an environmental strain to a host specific pathogen for CF patients.

5.3|Medical perspective

Understanding how bacterial populations adapt to a novel environment and which underlying genetic changes occur may provide useful information in the design of new treatment strategies and therapeutic targets.

The results presented here may therefore contribute to improve future therapeutic strategies since a clear understanding of the population diversity is important to target treatment. We show that there is high population diversity both contemporary and long-term and that distinct subpopulations are spatially distributed in the CF airway. Designing treatments that target the different site are therefore needed. In order to specify and target treatments properly it is necessary to know the precise localization and diagnosis of the CF airway colonizations and infections. Our current study suggest that subpopulations found in the lower airways at some point originated from the sinus population and migrated to the lung. The importance of sinuses as bacterial reservoirs has previously been indicated [66, 69]. To prevent or delay transition to chronic lung infection and ultimately prolong the life of the patients, early diagnosis and treatment of *P. aeruginosa* colonizations in the paranasal sinuses could be an important therapeutic approach [69].

With the realization that the intraclonal diversity is high, there need to be a better diagnosis and characterization of the strain variation especially in relation to antibiotics.

When a clinician needs to decide how to treat a bacterial infection, knowledge about diversity and its impact on phenotypic variation is important and can provide better treatments.

A clear understanding of within-host population dynamics during the course of infection and generation of diversity is therefore necessary for optimal therapeutic interventions.

References

1. Yang, L., et al., *In situ growth rates and biofilm development of Pseudomonas aeruginosa populations in chronic lung infections.* J Bacteriol, 2008. **190**(8): p. 2767-76.
2. Elena, S.F. and R.E. Lenski, *Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation.* Nat Rev Genet, 2003. **4**(6): p. 457-69.
3. Lenski, R.E., et al., *Evolution of competitive fitness in experimental populations of E. coli: what makes one genotype a better competitor than another?* Antonie van Leeuwenhoek, 1998. **73**(1): p. 35-47.
4. Sniegowski, P.D., P.J. Gerrish, and R.E. Lenski, *Evolution of high mutation rates in experimental populations of E. coli.* Nature, 1997. **387**(6634): p. 703-5.
5. Woods, R., et al., *Tests of parallel molecular evolution in a long-term experiment with Escherichia coli.* Proc Natl Acad Sci U S A, 2006. **103**(24): p. 9107-12.
6. Blount, Z.D., C.Z. Borland, and R.E. Lenski, *Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli.* Proc Natl Acad Sci U S A, 2008. **105**(23): p. 7899-906.
7. Ostrowski, E.A., D.E. Rozen, and R.E. Lenski, *Pleiotropic effects of beneficial mutations in Escherichia coli.* Evolution; international journal of organic evolution, 2005. **59**(11): p. 2343-52.
8. Leroi, A.M., A.F. Bennett, and R.E. Lenski, *Temperature acclimation and competitive fitness: an experimental test of the beneficial acclimation assumption.* Proc Natl Acad Sci U S A, 1994. **91**(5): p. 1917-21.
9. Dewitt, T.J., A. Sih, and D.S. Wilson, *Costs and limits of phenotypic plasticity.* Trends in ecology & evolution, 1998. **13**(2): p. 77-81.
10. Levin, B.R. and J.J. Bull, *Short-sighted evolution and the virulence of pathogenic microorganisms.* Trends in microbiology, 1994. **2**(3): p. 76-81.
11. Dettman, J.R., et al., *Evolutionary insight from whole-genome sequencing of experimentally evolved microbes.* Molecular ecology, 2012. **21**(9): p. 2058-77.
12. Gros, P.A., H. Le Nagard, and O. Tenaillon, *The evolution of epistasis and its links with genetic robustness, complexity and drift in a phenotypic model of adaptation.* Genetics, 2009. **182**(1): p. 277-93.
13. Desai, M.M. and D.S. Fisher, *Beneficial mutation selection balance and the effect of linkage on positive selection.* Genetics, 2007. **176**(3): p. 1759-98.
14. Barrick, J.E. and R.E. Lenski, *Genome-wide mutational diversity in an evolving population of Escherichia coli.* Cold Spring Harbor symposia on quantitative biology, 2009. **74**: p. 119-29.
15. Barrick, J.E. and R.E. Lenski, *Genome dynamics during experimental evolution.* Nat Rev Genet, 2013. **14**(12): p. 827-39.
16. Gerrish, P.J. and R.E. Lenski, *The fate of competing beneficial mutations in an asexual population.* Genetica, 1998. **102-103**(1-6): p. 127-44.
17. Bohannan, B.J., et al., *Trade-offs and coexistence in microbial microcosms.* Antonie van Leeuwenhoek, 2002. **81**(1-4): p. 107-15.
18. Lenski, R.E. and M. Travisano, *Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations.* Proc Natl Acad Sci U S A, 1994. **91**(15): p. 6808-14.
19. Elena, S.F. and R. Sanjuan, *Evolution. Climb every mountain?* Science, 2003. **302**(5653): p. 2074-5.
20. MacLean, R.C., G. Bell, and P.B. Rainey, *The evolution of a pleiotropic fitness tradeoff in Pseudomonas fluorescens.* Proc Natl Acad Sci U S A, 2004. **101**(21): p. 8072-7.
21. Kassen, R., *The experimental evolution of specialists, generalists, and the maintenance of diversity.* Journal of Evolutionary Biology, 2002. **15**(2): p. 173-190.
22. Rainey, P.B., et al., *The emergence and maintenance of diversity: insights from experimental bacterial populations.* Trends in ecology & evolution, 2000. **15**(6): p. 243-247.
23. Buckling, A., et al., *The Beagle in a bottle.* Nature, 2009. **457**(7231): p. 824-9.
24. Rainey, P.B. and M. Travisano, *Adaptive radiation in a heterogeneous environment.* Nature, 1998. **394**(6688): p. 69-72.

25. Boles, B.R., M. Thoendel, and P.K. Singh, *Self-generated diversity produces "insurance effects" in biofilm communities.* Proc Natl Acad Sci U S A, 2004. **101**(47): p. 16630-5.
26. Boles, B.R. and P.K. Singh, *Endogenous oxidative stress produces diversity and adaptability in biofilm communities.* Proc Natl Acad Sci U S A, 2008. **105**(34): p. 12503-8.
27. Markussen, T., et al., *Niche partitioning results in diversification and long-term co-existence of Pseudomonas aeruginosa sub-lineages during chronic CF airway infections.* Submitted, 2014.
28. Wang, L., F.F. Wang, and W. Qian, *Evolutionary rewiring and reprogramming of bacterial transcription regulation.* Journal of genetics and genomics = Yi chuan xue bao, 2011. **38**(7): p. 279-88.
29. Pujato, M., et al., *The underlying molecular and network level mechanisms in the evolution of robustness in gene regulatory networks.* PLoS Comput Biol, 2013. **9**(1): p. e1002865.
30. Hindre, T., et al., *New insights into bacterial adaptation through in vivo and in silico experimental evolution.* Nat Rev Microbiol, 2012. **10**(5): p. 352-65.
31. Cooper, T.F., et al., *Expression profiles reveal parallel evolution of epistatic interactions involving the CRP regulon in Escherichia coli.* PLoS genetics, 2008. **4**(2): p. e35.
32. Damkiaer, S., et al., *Evolutionary remodeling of global regulatory networks during long-term bacterial adaptation to human hosts.* Proc Natl Acad Sci U S A, 2013. **110**(19): p. 7766-71.
33. Markussen, T.M., et al., *Evolutionary modification of Pseudomonas aeruginosa Vfr functionality during chronic infection.* Submitted, 2014.
34. Oliver, A., et al., *High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection.* Science, 2000. **288**(5469): p. 1251-4.
35. Taddei, F., et al., *Role of mutator alleles in adaptive evolution.* Nature, 1997. **387**(6634): p. 700-2.
36. Tenaillon, O., et al., *Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria.* Genetics, 1999. **152**(2): p. 485-93.
37. Giraud, A., et al., *Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut.* Science, 2001. **291**(5513): p. 2606-8.
38. Tanaka, M.M., C.T. Bergstrom, and B.R. Levin, *The evolution of mutator genes in bacterial populations: the roles of environmental change and timing.* Genetics, 2003. **164**(3): p. 843-54.
39. Ciofu, O., et al., *Genetic adaptation of Pseudomonas aeruginosa during chronic lung infection of patients with cystic fibrosis: strong and weak mutators with heterogeneous genetic backgrounds emerge in mucA and/or lasR mutants.* Microbiology, 2010. **156**(Pt 4): p. 1108-19.
40. Ochman, H., J.G. Lawrence, and E.A. Groisman, *Lateral gene transfer and the nature of bacterial innovation.* Nature, 2000. **405**(6784): p. 299-304.
41. Ehrlich, G.D., et al., *Bacterial plurality as a general mechanism driving persistence in chronic infections.* Clinical orthopaedics and related research, 2005(437): p. 20-4.
42. Ehrlich, G.D., et al., *The distributed genome hypothesis as a rubric for understanding evolution in situ during chronic bacterial biofilm infectious processes.* FEMS immunology and medical microbiology, 2010. **59**(3): p. 269-79.
43. Gogarten, J.P. and J.P. Townsend, *Horizontal gene transfer, genome innovation and evolution.* Nat Rev Microbiol, 2005. **3**(9): p. 679-87.
44. Eisen, J.A., *Horizontal gene transfer among microbial genomes: new insights from complete genome analysis.* Current opinion in genetics & development, 2000. **10**(6): p. 606-11.
45. Lawrence, J.G. and H. Ochman, *Molecular archaeology of the Escherichia coli genome.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9413-7.
46. Markussen, T.M., et al., *Generation of genetic diversity among Pseudomonas aeruginosa via horizontal gene transfer during chronic CF infections.* Unpublished, 2014.
47. Hu, F.Z. and G.D. Ehrlich, *Population-level virulence factors amongst pathogenic bacteria: relation to infection outcome.* Future microbiology, 2008. **3**(1): p. 31-42.
48. Lee, B., et al., *Heterogeneity of biofilms formed by nonmucoid Pseudomonas aeruginosa isolates from patients with cystic fibrosis.* J Clin Microbiol, 2005. **43**(10): p. 5247-55.
49. Baur, B., et al., *Genetic transformation in freshwater: Escherichia coli is able to develop natural competence.* Appl Environ Microbiol, 1996. **62**(10): p. 3673-8.
50. Whitchurch, C.B., et al., *Extracellular DNA required for bacterial biofilm formation.* Science, 2002. **295**(5559): p. 1487.

51. Bernier, R. and H.H. Arsenault, *Deconvolution of two-dimensional images with zeros in the transfer function.* Applied optics, 1991. **30**(35): p. 5163-8.
52. Campana, S., et al., *Molecular epidemiology of Pseudomonas aeruginosa, Burkholderia cepacia complex and methicillin-resistant Staphylococcus aureus in a cystic fibrosis center.* J Cyst Fibros, 2004. **3**(3): p. 159-63.
53. Bear, C.E., et al., *Purification and functional reconstitution of the cystic fibrosis transmembrane conductance regulator (CFTR).* Cell, 1992. **68**(4): p. 809-18.
54. Riordan, J.R., et al., *Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA.* Science, 1989. **245**(4922): p. 1066-73.
55. Lyczak, J.B., C.L. Cannon, and G.B. Pier, *Lung infections associated with cystic fibrosis.* Clin Microbiol Rev, 2002. **15**(2): p. 194-222.
56. Foundation, C.F., *Cystic Fibrosis Foundation patient registry 2010 annual data report.* 2011.
57. Kerem, B., et al., *Identification of the cystic fibrosis gene: genetic analysis.* Science, 1989. **245**(4922): p. 1073-80.
58. Koch, C. and N. Hoiby, *Pathogenesis of cystic fibrosis.* Lancet, 1993. **341**(8852): p. 1065-9.
59. Du, K., M. Sharma, and G.L. Lukacs, *The DeltaF508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR.* Nat Struct Mol Biol, 2005. **12**(1): p. 17-25.
60. Boucher, R.C., *An overview of the pathogenesis of cystic fibrosis lung disease.* Adv Drug Deliv Rev, 2002. **54**(11): p. 1359-71.
61. Stammberger, H.R. and D.W. Kennedy, *Paranasal sinuses:anatomic terminology and nomenclature.* Ann Otol Rhinol Laryngol Suppl, 1995. **167**: p. 7-16.
62. Hoiby, N., *P. aeruginosa in Cystic Fibrosis Patients Resists Host Defenses, Antibiotics.* Microbe, 2006. **1**(12): p. 7.
63. Bjarnsholt, T., et al., *Pseudomonas aeruginosa biofilms in the respiratory tract of cystic fibrosis patients.* Pediatr Pulmonol, 2009. **44**(6): p. 547-58.
64. Hogg, J.C., *Pathophysiology of airflow limitation in chronic obstructive pulmonary disease.* Lancet, 2004. **364**(9435): p. 709-21.
65. Folkesson, A., et al., *Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective.* Nat Rev Microbiol, 2012. **10**(12): p. 841-51.
66. Mainz, J.G., et al., *Concordant genotype of upper and lower airways P aeruginosa and S aureus isolates in cystic fibrosis.* Thorax, 2009. **64**(6): p. 535-40.
67. Muhlebach, M.S., et al., *Are lower airway or throat cultures predictive of sinus bacteriology in cystic fibrosis?* Pediatr Pulmonol, 2006. **41**(5): p. 445-51.
68. Roby, B.B., et al., *Sinus surgery in cystic fibrosis patients: comparison of sinus and lower airway cultures.* Int J Pediatr Otorhinolaryngol, 2008. **72**(9): p. 1365-9.
69. Hansen, S.K., et al., *Evolution and diversification of Pseudomonas aeruginosa in the paranasal sinuses of cystic fibrosis children have implications for chronic lung infection.* ISME J, 2012. **6**(1): p. 31-45.
70. Johansen, H.K., et al., *Colonisation and infection of the paranasal sinuses in cystic fibrosis patients is accompanied by a reduced PMN response.* J Cyst Fibros, 2012. **11**(6): p. 525-31.
71. LiPuma, J.J., et al., *Person-to-person transmission of Pseudomonas cepacia between patients with cystic fibrosis.* Lancet, 1990. **336**(8723): p. 1094-6.
72. Salunkhe, P., et al., *A cystic fibrosis epidemic strain of Pseudomonas aeruginosa displays enhanced virulence and antimicrobial resistance.* J Bacteriol, 2005. **187**(14): p. 4908-20.
73. Yang, L., L. Jelsbak, and S. Molin, *Microbial ecology and adaptation in cystic fibrosis airways.* Environ Microbiol, 2011. **13**(7): p. 1682-9.
74. Hansen, C.R., T. Pressler, and N. Hoiby, *Early aggressive eradication therapy for intermittent Pseudomonas aeruginosa airway colonization in cystic fibrosis patients: 15 years experience.* J Cyst Fibros, 2008. **7**(6): p. 523-30.
75. Smith, E.E., et al., *Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients.* Proc Natl Acad Sci U S A, 2006. **103**(22): p. 8487-92.
76. Yang, L., et al., *Evolutionary dynamics of bacteria in a human host environment.* Proc Natl Acad Sci U S A, 2011. **108**(18): p. 7481-6.

77. Pier, G.B., *Pseudomonas aeruginosa lipopolysaccharide: a major virulence factor, initiator of inflammation and target for effective immunity.* International journal of medical microbiology : IJMM, 2007. **297**(5): p. 277-95.
78. Prince, A., *Flagellar activation of epithelial signaling.* American journal of respiratory cell and molecular biology, 2006. **34**(5): p. 548-51.
79. Palmer, K.L., L.M. Aye, and M. Whiteley, *Nutritional cues control Pseudomonas aeruginosa multicellular behavior in cystic fibrosis sputum.* J Bacteriol, 2007. **189**(22): p. 8079-87.
80. Barth, A.L. and T.L. Pitt, *The high amino-acid content of sputum from cystic fibrosis patients promotes growth of auxotrophic Pseudomonas aeruginosa.* Journal of medical microbiology, 1996. **45**(2): p. 110-9.
81. Boulette, M.L., et al., *Characterization of alanine catabolism in Pseudomonas aeruginosa and its importance for proliferation in vivo.* J Bacteriol, 2009. **191**(20): p. 6329-34.
82. Harrison, F., *Microbial ecology of the cystic fibrosis lung.* Microbiology, 2007. **153**(Pt 4): p. 917-23.
83. West, S.A., et al., *Social evolution theory for microorganisms.* Nat Rev Microbiol, 2006. **4**(8): p. 597-607.
84. Griffin, A.S., S.A. West, and A. Buckling, *Cooperation and competition in pathogenic bacteria.* Nature, 2004. **430**(7003): p. 1024-7.
85. Bragonzi, A., et al., *Modelling co-infection of the cystic fibrosis lung by Pseudomonas aeruginosa and Burkholderia cenocepacia reveals influences on biofilm formation and host response.* PLoS One, 2012. **7**(12): p. e52330.
86. Korgaonkar, A., et al., *Community surveillance enhances Pseudomonas aeruginosa virulence during polymicrobial infection.* Proc Natl Acad Sci U S A, 2013. **110**(3): p. 1059-64.
87. Ratjen, F., et al., *Diagnostic value of serum antibodies in early Pseudomonas aeruginosa infection in cystic fibrosis patients.* Pediatr Pulmonol, 2007. **42**(3): p. 249-55.
88. Aanaes, K., et al., *Clinical effects of sinus surgery and adjuvant therapy in cystic fibrosis patients - can chronic lung infections be postponed?* Rhinology, 2013. **51**(3): p. 222-30.
89. Pukatzki, S., R.H. Kessin, and J.J. Mekalanos, *The human pathogen Pseudomonas aeruginosa utilizes conserved virulence pathways to infect the social amoeba Dictyostelium discoideum.* Proc Natl Acad Sci U S A, 2002. **99**(5): p. 3159-64.
90. Rahme, L.G., et al., *Plants and animals share functionally common bacterial virulence factors.* Proc Natl Acad Sci U S A, 2000. **97**(16): p. 8815-21.
91. Apidianakis, Y. and L.G. Rahme, *Drosophila melanogaster as a model host for studying Pseudomonas aeruginosa infection.* Nat Protoc, 2009. **4**(9): p. 1285-94.
92. Tan, M.W., S. Mahajan-Miklos, and F.M. Ausubel, *Killing of Caenorhabditis elegans by Pseudomonas aeruginosa used to model mammalian bacterial pathogenesis.* Proc Natl Acad Sci U S A, 1999. **96**(2): p. 715-20.
93. Clatworthy, A.E., et al., *Pseudomonas aeruginosa infection of zebrafish involves both host and pathogen determinants.* Infect Immun, 2009. **77**(4): p. 1293-303.
94. Miyata, S., et al., *Use of the Galleria mellonella caterpillar as a model host to study the role of the type III secretion system in Pseudomonas aeruginosa pathogenesis.* Infect Immun, 2003. **71**(5): p. 2404-13.
95. Sadikot, R.T., et al., *Pathogen-host interactions in Pseudomonas aeruginosa pneumonia.* American journal of respiratory and critical care medicine, 2005. **171**(11): p. 1209-23.
96. Santucci, S.G., et al., *Infections in a burn intensive care unit: experience of seven years.* The Journal of hospital infection, 2003. **53**(1): p. 6-13.
97. Stover, C.K., et al., *Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen.* Nature, 2000. **406**(6799): p. 959-64.
98. Callaghan, M. and S. McClean, *Bacterial host interactions in cystic fibrosis.* Curr Opin Microbiol, 2012. **15**(1): p. 71-7.
99. Dekimpe, V. and E. Deziel, *Revisiting the quorum-sensing hierarchy in Pseudomonas aeruginosa: the transcriptional regulator RhlR regulates LasR-specific factors.* Microbiology, 2009. **155**(Pt 3): p. 712-23.
100. Hauser, A.R. and E.A. Ozer, *Pseudomonas aeruginosa.* Nat Rev Microbiol, 2011. **9 (Poster)**.

101. Pesci, E.C., et al., *Regulation of las and rhl quorum sensing in Pseudomonas aeruginosa.* J Bacteriol, 1997. **179**(10): p. 3127-32.
102. Diggle, S.P., et al., *4-quinolone signalling in Pseudomonas aeruginosa: old molecules, new perspectives.* International journal of medical microbiology : IJMM, 2006. **296**(2-3): p. 83-91.
103. Pesci, E.C., et al., *Quinolone signaling in the cell-to-cell communication system of Pseudomonas aeruginosa.* Proc Natl Acad Sci U S A, 1999. **96**(20): p. 11229-34.
104. Gambello, M.J. and B.H. Iglewski, *Cloning and characterization of the Pseudomonas aeruginosa lasR gene, a transcriptional activator of elastase expression.* J Bacteriol, 1991. **173**(9): p. 3000-9.
105. Pearson, J.P., et al., *A second N-acylhomoserine lactone signal produced by Pseudomonas aeruginosa.* Proc Natl Acad Sci U S A, 1995. **92**(5): p. 1490-4.
106. Latifi, A., et al., *A hierarchical quorum-sensing cascade in Pseudomonas aeruginosa links the transcriptional activators LasR and RhIR (VsmR) to expression of the stationary-phase sigma factor RpoS.* Mol Microbiol, 1996. **21**(6): p. 1137-46.
107. Jimenez, P.N., et al., *The multiple signaling systems regulating virulence in Pseudomonas aeruginosa.* Microbiol Mol Biol Rev, 2012. **76**(1): p. 46-65.
108. Lazdunski, A.M., I. Ventre, and J.N. Sturgis, *Regulatory circuits and communication in Gram-negative bacteria.* Nat Rev Microbiol, 2004. **2**(7): p. 581-92.
109. Venturi, V., *Regulation of quorum sensing in Pseudomonas.* FEMS Microbiol Rev, 2006. **30**(2): p. 274-91.
110. Suh, S.J., et al., *Effect of vfr mutation on global gene expression and catabolite repression control of Pseudomonas aeruginosa.* Microbiology, 2002. **148**(Pt 5): p. 1561-9.
111. VanderLeest, H.T., et al., *Seasonal encoding by the circadian pacemaker of the SCN.* Curr Biol, 2007. **17**(5): p. 468-73.
112. Albus, A.M., et al., *Vfr controls quorum sensing in Pseudomonas aeruginosa.* J Bacteriol, 1997. **179**(12): p. 3928-35.
113. Croda-Garcia, G., et al., *Transcriptional regulation of Pseudomonas aeruginosa rhlR: role of the CRP orthologue Vfr (virulence factor regulator) and quorum-sensing regulators LasR and RhlR.* Microbiology, 2011. **157**(Pt 9): p. 2545-55.
114. Whitchurch, C.B., et al., *Pseudomonas aeruginosa fimL regulates multiple virulence functions by intersecting with Vfr-modulated pathways.* Mol Microbiol, 2005. **55**(5): p. 1357-78.
115. Lam, J.S., et al., *Genetic and Functional Diversity of Pseudomonas aeruginosa Lipopolysaccharide.* Front Microbiol, 2011. **2**: p. 118.
116. King, J.D., et al., *Review: Lipopolysaccharide biosynthesis in Pseudomonas aeruginosa.* Innate Immun, 2009. **15**(5): p. 261-312.
117. Kocincova, D. and J.S. Lam, *Structural diversity of the core oligosaccharide domain of Pseudomonas aeruginosa lipopolysaccharide.* Biochemistry (Mosc), 2011. **76**(7): p. 755-60.
118. Rivera, M., et al., *Heterogeneity of lipopolysaccharides from Pseudomonas aeruginosa: analysis of lipopolysaccharide chain length.* J Bacteriol, 1988. **170**(2): p. 512-21.
119. Rivera, M. and E.J. McGroarty, *Analysis of a common-antigen lipopolysaccharide from Pseudomonas aeruginosa.* J Bacteriol, 1989. **171**(4): p. 2244-8.
120. Dasgupta, T., et al., *Characterization of lipopolysaccharide-deficient mutants of Pseudomonas aeruginosa derived from serotypes O3, O5, and O6.* Infect Immun, 1994. **62**(3): p. 809-17.
121. Fomsgaard, A., M.A. Freudenberg, and C. Galanos, *Modification of the silver staining technique to detect lipopolysaccharide in polyacrylamide gels.* J Clin Microbiol, 1990. **28**(12): p. 2627-31.
122. Lam, M.Y., et al., *Occurrence of a common lipopolysaccharide antigen in standard and clinical strains of Pseudomonas aeruginosa.* J Clin Microbiol, 1989. **27**(5): p. 962-7.
123. Hancock, R.E., et al., *Pseudomonas aeruginosa isolates from patients with cystic fibrosis: a class of serum-sensitive, nontypable strains deficient in lipopolysaccharide O side chains.* Infect Immun, 1983. **42**(1): p. 170-7.
124. Penketh, A., et al., *The relationship of phenotype changes in Pseudomonas aeruginosa to the clinical condition of patients with cystic fibrosis.* The American review of respiratory disease, 1983. **127**(5): p. 605-8.

125. Kadurugamuwa, J.L., J.S. Lam, and T.J. Beveridge, *Interaction of gentamicin with the A band and B band lipopolysaccharides of Pseudomonas aeruginosa and its possible lethal effect.* Antimicrob Agents Chemother, 1993. **37**(4): p. 715-21.
126. Rivera, M., et al., *Enhanced binding of polycationic antibiotics to lipopolysaccharide from an aminoglycoside-supersusceptible, tolA mutant strain of Pseudomonas aeruginosa.* Antimicrob Agents Chemother, 1988. **32**(5): p. 649-55.
127. Kohler, T., V. Donner, and C. van Delden, *Lipopolysaccharide as shield and receptor for R-pyocin-mediated killing in Pseudomonas aeruginosa.* J Bacteriol, 2010. **192**(7): p. 1921-8.
128. Meadow, P.M. and P.L. Wells, *Receptor sites for R-type pyocins and bacteriophage E79 in the core part of the lipopolysaccharide of Pseudomonas aeruginosa PAC1.* J Gen Microbiol, 1978. **108**: p. 339-343.
129. Kocincova, D. and J.S. Lam, *A deletion in the wapB promoter in many serotypes of Pseudomonas aeruginosa accounts for the lack of a terminal glucose residue in the core oligosaccharide and resistance to killing by R3-pyocin.* Mol Microbiol, 2013. **89**(3): p. 464-78.
130. Michel-Briand, Y. and C. Baysse, *The pyocins of Pseudomonas aeruginosa.* Biochimie, 2002. **84**(5-6): p. 499-510.
131. Williams, S.R., et al., *Retargeting R-type pyocins to generate novel bactericidal protein complexes.* Appl Environ Microbiol, 2008. **74**(12): p. 3868-76.
132. Riley, M.A. and J.E. Wertz, *Bacteriocins: evolution, ecology, and application.* Annual review of microbiology, 2002. **56**: p. 117-37.
133. Ernst, R.K., et al., *Specific lipopolysaccharide found in cystic fibrosis airway Pseudomonas aeruginosa.* Science, 1999. **286**(5444): p. 1561-5.
134. Jensen, E.T., et al., *Epidemiology of Pseudomonas aeruginosa in cystic fibrosis and the possible role of contamination by dental equipment.* The Journal of hospital infection, 1997. **36**(2): p. 117-22.
135. Rau, M.H., et al., *Early adaptive developments of Pseudomonas aeruginosa after the transition from life in the environment to persistent colonization in the airways of human cystic fibrosis hosts.* Environ Microbiol, 2010. **12**(6): p. 1643-58.
136. Moreau-Marquis, S., B.A. Stanton, and G.A. O'Toole, *Pseudomonas aeruginosa biofilm formation in the cystic fibrosis airway.* Pulmonary pharmacology & therapeutics, 2008. **21**(4): p. 595-9.
137. Landry, R.M., et al., *Mucin-Pseudomonas aeruginosa interactions promote biofilm formation and antibiotic resistance.* Mol Microbiol, 2006. **59**(1): p. 142-51.
138. Hansen, S.K., et al., *Evolution of species interactions in a biofilm community.* Nature, 2007. **445**(7127): p. 533-6.
139. Hancock, R.E. and D.P. Speert, *Antibiotic resistance in Pseudomonas aeruginosa: mechanisms and impact on treatment.* Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy, 2000. **3**(4): p. 247-255.
140. Deretic, V., M.J. Schurr, and H. Yu, *Pseudomonas aeruginosa, mucoidy and the chronic infection phenotype in cystic fibrosis.* Trends in microbiology, 1995. **3**(9): p. 351-6.
141. Haussler, S., et al., *Small-colony variants of Pseudomonas aeruginosa in cystic fibrosis.* Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 1999. **29**(3): p. 621-5.
142. D'Argenio, D.A., et al., *Growth phenotypes of Pseudomonas aeruginosa lasR mutants adapted to the airways of cystic fibrosis patients.* Mol Microbiol, 2007. **64**(2): p. 512-33.
143. Huse, H.K., et al., *Parallel evolution in Pseudomonas aeruginosa over 39,000 generations in vivo.* mBio, 2010. **1**(4).
144. Ashish, A., et al., *Extensive diversification is a common feature of Pseudomonas aeruginosa populations during respiratory infections in cystic fibrosis.* J Cyst Fibros, 2013. **12**(6): p. 790-3.
145. Mowat, E., et al., *Pseudomonas aeruginosa population diversity and turnover in cystic fibrosis chronic infections.* American journal of respiratory and critical care medicine, 2011. **183**(12): p. 1674-9.
146. Workentine, M.L., et al., *Phenotypic heterogeneity of Pseudomonas aeruginosa populations in a cystic fibrosis patient.* PLoS One, 2013. **8**(4): p. e60225.
147. Bernier, S.P., M.L. Workentine, and M.G. Surette, *Genetic signature of bacterial pathogen adaptation during chronic pulmonary infections.* Nature genetics, 2014. **46**(1): p. 5-6.

148. Martin, D.W., et al., *Mechanism of conversion to mucoidy in Pseudomonas aeruginosa infecting cystic fibrosis patients.* Proc Natl Acad Sci U S A, 1993. **90**(18): p. 8377-81.
149. Oliver, A.M. and D.M. Weir, *Inhibition of bacterial binding to mouse macrophages by Pseudomonas alginate.* Journal of clinical & laboratory immunology, 1983. **10**(4): p. 221-4.
150. Schwarzmann, S. and J.R. Boring, *Antiphagocytic Effect of Slime from a Mucoid Strain of Pseudomonas aeruginosa.* Infect Immun, 1971. **3**(6): p. 762-7.
151. Simpson, J.A., S.E. Smith, and R.T. Dean, *Scavenging by alginate of free radicals released by macrophages.* Free radical biology & medicine, 1989. **6**(4): p. 347-53.
152. Boucher, J.C., et al., *Mucoid Pseudomonas aeruginosa in cystic fibrosis: characterization of muc mutations in clinical isolates and analysis of clearance in a mouse model of respiratory infection.* Infect Immun, 1997. **65**(9): p. 3838-46.
153. Vogne, C., et al., *Role of the multidrug efflux system MexXY in the emergence of moderate resistance to aminoglycosides among Pseudomonas aeruginosa isolates from patients with cystic fibrosis.* Antimicrob Agents Chemother, 2004. **48**(5): p. 1676-80.
154. Cramer, N., et al., *Microevolution of the major common Pseudomonas aeruginosa clones C and PA14 in cystic fibrosis lungs.* Environ Microbiol, 2011. **13**(7): p. 1690-704.
155. Nguyen, D. and P.K. Singh, *Evolving stealth: genetic adaptation of Pseudomonas aeruginosa during cystic fibrosis infections.* Proc Natl Acad Sci U S A, 2006. **103**(22): p. 8305-6.
156. Lieberman, T.D., et al., *Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures.* Nature genetics, 2014. **46**(1): p. 82-7.
157. Yachi, S. and M. Loreau, *Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis.* Proc Natl Acad Sci U S A, 1999. **96**(4): p. 1463-8.
158. Foweraker, J.E., et al., *Phenotypic variability of Pseudomonas aeruginosa in sputa from patients with acute infective exacerbation of cystic fibrosis and its impact on the validity of antimicrobial susceptibility testing.* The Journal of antimicrobial chemotherapy, 2005. **55**(6): p. 921-7.
159. Chung, J.C., et al., *Genomic variation among contemporary Pseudomonas aeruginosa isolates from chronically infected cystic fibrosis patients.* J Bacteriol, 2012. **194**(18): p. 4857-66.
160. Hoiby, N., *Pseudomonas aeruginosa infection in cystic fibrosis. Diagnostic and prognostic significance of pseudomonas aeruginosa precipitins determined by means of crossed immunoelectrophoresis. A survey.* Acta Pathol Microbiol Scand Suppl, 1977(262): p. 1-96.
161. Jelsbak, L., et al., *Molecular epidemiology and dynamics of Pseudomonas aeruginosa populations in lungs of cystic fibrosis patients.* Infect Immun, 2007. **75**(5): p. 2214-24.
162. Pedersen, S.D., *Systems Biology Investigations of Pseudomonas aeruginosa evolution in association with Human airway infections (PhD Thesis)*, in *Department of Systems Biology* 2013, Technical University of Denmark.
163. Marvig, R.L., et al., *Genome Analysis of a Transmissible Lineage of Pseudomonas aeruginosa Reveals Pathoadaptive Mutations and Distinct Evolutionary Paths of Hypermutators.* PLoS genetics, 2013. **9**(9): p. e1003741.
164. Barrick, J.E., et al., *Genome evolution and adaptation in a long-term experiment with Escherichia coli.* Nature, 2009. **461**(7268): p. 1243-7.
165. Hoiby, N., et al., *Pseudomonas aeruginosa and the in vitro and in vivo biofilm mode of growth.* Microbes and infection / Institut Pasteur, 2001. **3**(1): p. 23-35.
166. Coutinho, C.P., et al., *Long-term colonization of the cystic fibrosis lung by Burkholderia cepacia complex bacteria: epidemiology, clonal variation, and genome-wide expression alterations.* Frontiers in cellular and infection microbiology, 2011. **1**: p. 12.
167. Schott, T., et al., *Microevolution of a zoonotic Helicobacter population colonizing the stomach of a human host before and after failed treatment.* Genome biology and evolution, 2012. **4**(12): p. 1310-5.
168. Wong, A., N. Rodrigue, and R. Kassen, *Genomics of adaptation during experimental evolution of the opportunistic pathogen Pseudomonas aeruginosa.* PLoS genetics, 2012. **8**(9): p. e1002928.
169. Cooper, V.S. and R.E. Lenski, *The population genetics of ecological specialization in evolving Escherichia coli populations.* Nature, 2000. **407**(6805): p. 736-9.
170. Ensminger, A.W., *Experimental evolution of pathogenesis: "patient" research.* PLoS Pathog, 2013. **9**(5): p. e1003340.

171. Lieberman, T.D., et al., *Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes.* Nature genetics, 2011. **43**(12): p. 1275-80.

Chapter 6

Research papers

The research papers included in this thesis are enclosed in the following order:

Paper 1

Markussen, T., Marvig, R. L., Gómez-Lozano, M., Burleigh, A. E., Høiby, N., Johansen, H. K., Molin, S., and Jelsbak, J. (2014) Niche partitioning results in diversification and long-term co-existence of distinct *Pseudomonas aeruginosa* sub-lineages during chronic CF airway infections. *Manuscript submitted*

Paper 2

Markussen, T. M. M., Nielsen, T., Molin, S., and Jelsbak, L. (2014) Evolutionary modification of *Pseudomonas aeruginosa* Vfr functionality during chronic infection. *Manuscript submitted.*

Paper 3

Markussen, T. M. M, Marvig, R. L., Johansen, H. K., Lam, J. S., and Jelsbak, L. (2014) Generation of genetic diversity among Pseudomonas aeruginosa via horizontal gene transfer during chronic CF infections. *Manuscript in preparation.*

Niche partitioning results in diversification and long-term co-existence of *Pseudomonas aeruginosa* sub-lineages during chronic CF airway infections

Trine Markussen[1], Rasmus L. Marvig[1], María Gómez-Lozano[1], Kasper Aanæs[4], Alexandra E. Burleigh[1], Niels Høiby[2,3], Helle Krogh Johansen[2], Søren Molin[1*], and Lars Jelsbak[1*]

[1]Department of Systems Biology, Technical University of Denmark, 2800, Lyngby, Denmark

[2]Department of Clinical Microbiology, University Hospital, Rigshospitalet, 2100 Copenhagen

[3]Institute for International Health, Immunology and Microbiology, University of Copenhagen, 2200 Copenhagen, Denmark

[4]Department of Otorhinolaryngology, Head and Neck Surgery, Rigshospitalet and Faculty of Health Sciences Copenhagen, Copenhagen, Denmark

* Corresponding authors: Lars Jelsbak and Søren Molin, Department of Systems Biology, Technical University of Denmark, 2800 Lyngby, Denmark. Email: sm@bio.dtu.dk (Søren Molin) and lj@bio.dtu.dk (Lars Jelsbak). Telephone: +45 45256129.

Classification: Biological Sciences, Microbiology.

Keywords: Pathogen evolution, adaptive radiation, chronic infection, cystic fibrosis, population diversity

Abstract

The opportunistic pathogen *Pseudomonas aeruginosa* chronically infects the airways of patients with cystic fibrosis (CF). A clear understanding of the within-host population dynamics during the course of infection is necessary for optimal therapeutic interventions. While *P. aeruginosa* population diversity within infected individuals has been observed by inspection of bacteria cultured from the same respiratory specimens, our understanding of how this diversity is generated and maintained is limited. To gain new insight into the bacterial diversification processes, we examined the temporal and spatial variations in the *P. aeruginosa* population diversity in a chronically infected CF patient during a 32-year period. Genome sequence data and phenotypic profiling of clonal bacterial isolates sampled from different infection sites demonstrated that multiple pathways for adaptation existed for the ancestral genotype, and that the original infecting clone rapidly diversified into three distinct sub-lineages, each with their own functional and genomic signatures and rates of adaptation. We further show that the three sub-lineages co-existed in the infected host for decades. We hypothesized that the observed diversification and stable maintenance of population polymorphisms was the result of partitioning of the sub-lineages into different niches. Indeed, sampling of the paranasal sinuses and the lower airways clearly demonstrated that the sub-lineages occupied separate geographical niches within the infected airways. The results point toward an underlying complex population structure not previously realized, and show that the highly structured CF host environment can contribute significantly to generation of intraclonal population diversity and its long-term maintenance.

Significance statement

Within-host pathogen evolution and diversification during the course of chronic infections is of importance in relation to therapeutic intervention strategies, yet our understanding of these processes is limited. Here we use chronic colonization of cystic fibrosis airways by the opportunistic pathogen *Pseudomonas aeruginosa* as a system for studying how evolution shape pathogen populations within hosts. We analyse clonal bacterial isolates sampled during a 32 year period from separate geographical niches within infected CF airways. We show that the original infecting strain immediately after initial colonization diverged into distinct sub-lineages that co-existed for decades, and occupied distinct niches. These results suggest that the spatial heterogeneity in CF airways plays a major role in relation to generation and stability of population diversity.

Introduction

Cystic fibrosis (CF) is a life-shortening disease caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. Although CF itself is a genetic disease, the morbidity and mortality is the result of life-long, chronic bacterial airway infections and the resulting inflammation. The opportunistic pathogen *Pseudomonas aeruginosa* is a predominant pathogen in relation to the chronic airway infections. CF patients often acquire intermittent colonization of the airways from early childhood, which eventually proceeds into chronic infection where the same clonal *P. aeruginosa* lineage can persist for decades despite antibiotic therapy (1). Once the chronic infection has been established, lineage replacement is rare (2). CF airway infections are associated with genetic and phenotypic changes in *P. aeruginosa*, which contribute to its long-term persistence in the airways. It is well established that the phenotype characteristics of *P. aeruginosa* isolates from chronically infected CF patients are most often remarkably different from the phenotypes displayed by the clonal isolates that initiated the infection in the same patients years before. Many phenotype characteristics of chronically infecting strains are consistently selected in different CF patients, suggesting that adaptation occurs with conserved patterns of evolution (3). This conclusion has been further substantiated by genome sequence analyses, which have led to the discovery of recurrent patterns of mutations in independent evolving *P. aeruginosa* lineages, and have furthermore provided new insight into the genetic changes underlying the phenotypic changes (4-6).

In parallel to these advances in characterizing the evolutionary pathways in *P. aeruginosa*, other studies have documented the existence of *P. aeruginosa* population diversity within individual hosts. For example, a significant level of phenotypic variability within the infecting *P. aeruginosa* population is commonly

observed, and *P. aeruginosa* clones sampled from the same clinical specimen often display large differences in colony morphology, quorum sensing regulation, antibiotic resistance and other phenotypes typically associated with chronic CF infections (7-11). Similarly, genome studies have provided evidence for genomic diversity among clonal isolates (12) and for clonal sub-lineage replacements during long-term infections in individual patients (4). Similar results have been obtained for other CF-associated pathogens (13). While these studies add to an emerging picture of an underlying and complex population dynamics within the infected host, the specific causes of the observed intraclonal diversity are difficult to unravel and are not well understood. Since the CF airways is a highly structured environment with multiple connected compartments each with different environmental conditions, a possible model is that population diversity is the result of evolution of sub-lineages that can efficiently occupy different niches available in the CF airways. Nevertheless, other processes such as interactions between variant sub-populations (*e.g.* evolution of syntropic relationships) (14), and the evolution of variants with increased mutation rates (*i.e.* the hypermutator phenotype) (15) have been described for *P. aeruginosa* CF isolates and could also contribute to population diversification.

Critically, it remains unknown how the observed intraclonal diversity is generated and if the environmental heterogeneity and structure of the CF airways contribute to diversity generation, how stable these variant populations are as a function of time, and whether the variant populations are geographically segregated in the airways. To address these issues, we analyzed clonal bacterial isolates sampled from a chronically infected individual during a 32-year period. Sampling and storage of *P. aeruginosa* isolates from chronically infected Danish CF patients has been carried out since 1973, which has resulted in a strain collection that represents 40 years of 'infection history' from a large number of infected individuals. These frozen stocks of *P. aeruginosa* can easily be revived and analyzed. In our initial examination of the clonal relationship among the stored isolates using low-resolution molecular typing methods we identified two distinct and dominant *P. aeruginosa* lineages that were observed repeatedly because of transmission among patients (16). Here, we focus on one of these transmissible lineages called DK1 and demonstrate rapid diversification of the original infecting strain into three distinct sub-lineages, each with their own functional and genomic signatures. We show that this functional and genomic diversification remained stable for decades, and demonstrate that the sub-lineages occupy separate niches such as the sinuses and the lower airways.

Results

Genomic evolution of the DK1 lineage and within-patient diversity. The DK1 lineage has been isolated from more than 40 CF patients since the start of the sampling program in 1973. We first sought to identify patients with a long clonal infection history with DK1, and from whom both contemporary DK1 isolates as well as historical isolates from lung expectorates were available. We found one such patient (P30M0), and sequenced the genomes of 18 isolates sampled between 1980 and 2012 from this patient (Figure 1). Furthermore, to characterize the evolution of the DK1 lineage from its entry into the CF environment until present day, we also identified and sequenced the four oldest DK1 isolates in the strain collection. These four isolates were sampled from four different patients in 1973 (Figure 1).

Following sequencing, we identified high-quality single nucleotide polymorphisms (SNPs) in the non-repetitive parts of the genomes by mapping sequence reads for each isolate against the DK1-P33F0-1973 genome sequence as previously described (4), and found a total of 160 SNPs that had appeared since the isolates diverged from their common DK1 ancestor before 1973.

The identification of SNPs enabled us to determine the evolutionary relationship among the DK1 clones (Figure 2A). Focusing on the isolates from P30M0, we observed that relatively few SNP changes had occurred during 32 years of colonization of the patient (Table S1). For example, only 19 SNP changes separate isolates DK1-P30M0-1980 and DK1-P30M0-2012c, and DK1-P30M0-2012c had only accumulated 24 SNP changes in total since it diverged from the common ancestor of all DK1 isolates (Figure 2A). Despite this moderate genomic evolution, the phylogenetic tree revealed a divergent population structure of the P30M0 isolates, and the deeply branched topology of the tree provided evidence for three distinct, coexisting sub-lineages (cluster A, B, and C in Figure 2A). We performed a Bayesian analysis to estimate the divergence dates and mutation rate. Based on the Bayesian phylogenetic reconstruction, we estimated that the infecting population first diverged in 1979 (95% highest posterior density (HPD): 1976-1982) (Figure 2A). This estimate coincides with the first DK1 culture from P30M0 which was sampled in 1980, suggesting that the DK1 population diverged soon after initial colonization. The population further diverged into cluster A and C a decade later (Figure 2A). Importantly, bacterial isolates from the three clusters were identified in contemporary samples (from 2012/2013) demonstrating their co-existence for decades in the same CF patient (see below).

Overall, the DK1 lineage accumulated 1.3 SNPs/year, and the relative rates of nonsynonymous and synonymous substitutions (dN/dS=0.56) implied that the long-term evolution of the DK1 lineage was dominated by negative selection ($P < 4.72\times10^{-10}$). Nonetheless, isolates from cluster C had a higher

evolutionary rate than the two other clusters (Figure 2B). The mean mutation rate for cluster C was estimated to be 2.7 SNPs/year, while the rates for cluster A and B were found to be 0.9 and 1.2 SNPs/year, respectively. The elevated rate in cluster C was not a consequence of horizontally acquired SNPs as observed in other CF isolates (4) or a hypermutator phenotype caused by mutations in DNA mismatch repair systems (17). Instead, we speculated that the elevated mutation rate was due to selection and fixation of adaptive mutations. In support of this hypothesis, we found a stronger signature of positive selection (*i.e.* a larger fraction of nonsynonymous mutations) among the SNPs in the C cluster (branches PD, PF, PG, XB, and XC in Figure 2A) relative to the A and B cluster (39/13 vs. 37/29; Fisher's exact test, P = 0.026).

We have previously identified a set of 65 pathoadaptive genes in *P. aeruginosa* involved in CF host adaptation (*i.e.* genes that undergo adaptive evolution under the pressure of natural selection during chronic CF infections) (4). We reasoned that the prevalence of pathoadaptive mutations in each of the three sub-lineages would provide an independent measurement of the level of positively selected SNPs. In agreement with our finding of enhanced positive selection in cluster C, we found that most pathoadaptive mutations were found in this cluster, whereas cluster A contained fewest (Figure 2C).

Phenotype analysis supports the existence of different sub-lineages. We hypothesised that the existence of three divergent sub-lineages each with different evolutionary trajectories would result in observable functional diversification. We first analysed the genotype/phenotype relationship within each cluster by measuring the catabolic performance of the isolates on 125 informative substrates of both carbon and nitrogen sources using Biolog Phenotype Microarrays. Hierarchical clustering (Ward's Linkage, Euclidian distance) of the data, resulted in four principal clusters, each containing isolates with related catabolic profiles (Figure 3C). It was observed that early isolates (from 1973) clustered together with *P. aeruginosa* laboratory reference strain PAO1, and that the later isolates from P30M0 grouped in three different clusters identical to the clusters derived from the genome sequence data (cf. Figure 2A).

P. aeruginosa isolates from chronically infected patients often display common phenotypic traits including conversion into alginate overproducing mucoid variants (18, 19), reduced growth rates (20), and loss of virulence factor production (5), quorum sensing (21) and motility (22). We measured variations in these phenotypes for isolates from the three clusters. All isolates had reduced or abolished motility, but lineage-specific differences in other traits were observed (Figure 3A). For example, cluster B isolates had maintained production of quorum-sensing (QS) molecules and had reduced protease secretion, while isolates from cluster A produced QS molecules but no proteases. Cluster C isolates had lost both QS production and protease secretion (Figure 3A). Furthermore isolates from this cluster had a significantly

longer *in vitro* doubling time than isolates from cluster A and B (Figure 3B). All P30M0 isolates contain a nonsense mutation (C349T) in the anti-sigma factor *mucA*. Inactivation of MucA leads to activation of the AlgT sigma factor, expression of the *alg* operon (encoding enzymes for alginate biosynthesis), and overproduction of the alginate exopolysaccharide. As expected from the *mucA* mutation, isolates from cluster A and C were mucoid. However, isolates from cluster B were non-mucoid most likely due to missense mutations in *algT* (Table S1).

Overall, the unique phenotypic profiles in relation to both catabolic capacities and CF-related phenotypes found for each cluster supported the genomic SNP-based evidence for co-existing sub-populations. Furthermore, we note that the isolates from cluster C showed more phenotype traits typically associated with chronic infections than the other two clusters.

Correlation between genomic evolution and phenotype diversity. Long-term colonization of CF airways is often associated with significant changes in global gene expression profiles of the infecting isolates (23, 24), and to examine the level of diversity and its variation as a function of time within each of the three clusters, we measured whole-genome transcript abundance on two temporally separated isolates from each cluster. The two isolates (DK1-P30M0-2001b and DK1-P30M0-2007b) from cluster A showed very similar gene expression profiles as indicated by the high Pearson correlation (r = 0.9944) (Figure 4A) and their close clustering in singular value decomposition (SVD) analysis shown in Figure 4D. In fact, only two genes with significant expression changes (Student's t-test, $P < 0.05$) were identified in isolate DK1-P30M0-2001b relative to DK1-P30M0-2007b. In contrast, temporally separated isolates from both cluster C and B showed a more diverse gene expression profile (r = 0.9632 and r= 0.9450, respectively). 78 genes with significant expression changes were identified when comparing isolates from cluster C (Figure 4C) and 298 genes with significant expression changes were identified when comparing the two non-mucoid isolates DK1-P30M0-2001a and DK1-P30M0-2009 from cluster B (Figure 4B). The differences in the extent of phenotypic diversity developing over time within the three clusters is consistent with a hypothesis that isolates from cluster B and C evolved in environments different from cluster A with stronger selective pressures mediating higher levels of both genomic evolution and phenotypic diversity.

Limited parallel evolution among coexisting sub-lineages. The long-term coexistence of distinct sub-lineages enabled us to investigate the extent of parallelism among three evolutionary pathways within the same individual. We examined the genome data for mutations that accumulated independently in the different sub-lineages (Table S1). As anticipated such parallel evolution was observed in genes known to be involved in antibiotic resistance, and included mutations in DNA gyrases *gyrA* (cluster B and cluster C), *gyrB* (cluster A and C), as well as *fusA1* (cluster B and C) encoding elongation factor G (25, 26) (Table S1).

Unexpectedly, no other examples of parallel genomic evolution were found which suggests that the three sub-lineages exist in separate niches each with different selective conditions by which evolution is directed. Importantly, this conclusion is in agreement with the observed lineage-specific distribution of pathoadaptive mutations (Figure 2C).

Mapping co-existing sub-population to different spatial sites. Patient P30M0 went through paranasal sinus surgery in 2012 during which expectorated sputum, bronchoalveolar lavage (BAL) and all sinus material (from both left and right sinuses) was sampled and stored. This provided a unique opportunity to determine the distribution of the three sub-lineages at different sites in the infected CF airways (sinus versus lower airways). Sub-lineage-specific primers were designed on the basis of the genome sequence data, and a cross-sectional analysis of multiple bacterial isolates from each sample was done by PCR screening (Materials and Methods). We found that all analysed isolates from the sinus samples belonged to cluster A (64 isolates in total), while all isolates from the BAL sample belonged to cluster C (64 isolates in total). Both clusters could be detected in the sputum sample as well as in additional sputum samples available from 2010, 2011 and 2012. However, in these samples cluster A isolates were found only in low numbers and constituted less than 2% of the total population. Bacterial isolates from cluster B were not found in the samples related to the upper airways (sinus surgery), and were infrequently detected in sputum samples from 2010 to 2013, where they were present at low numbers (less than 1%). Nevertheless, cluster B isolates had previously been consistently detected in sputum samples from the patient. Genome sequencing of one isolate from the BAL sample and two isolates sampled from each side of the sinuses verified that they belonged to cluster C (BAL isolate DK1-P30M0-2012a) and cluster A (sinus isolates DK1-P30M0-2012b and DK1-P30M0-2012c). Together, these findings confirmed that the three co-existing sub-populations evolved in spatially separated niches of the CF airway: isolates from cluster A colonizing the sinuses of this patient, and isolates from cluster B and C colonizing and evolving in the lower airways.

Discussion.

Although *P. aeruginosa* population heterogeneity in relation to chronic CF infections has frequently been documented by microbial culturing, our understanding of the temporal and spatial dynamics of this diversity is limited. In this study we combined current genome sequencing technologies, phenotypic profiling and unique sampling materials which included clonal bacterial isolates sampled during a 32-year period from both sputum and from material obtained from paranasal sinus surgery in a chronically infected CF patient to gain new insight into the bacterial diversification processes in chronic CF infections.

The phylogenomic analysis demonstrated that the initial infecting DK1 strain diversified into three sub-

populations (clusters A, B and C) each with their individual genetic signature of mutations and limited parallel genomic evolution. Our phenotype profiling of strains from the different clusters supported this result and showed that the accumulation of mutations shaped their phenotype in such ways that the three clusters were clearly distinguishable with respect to mucoidy, doubling time, metabolic performance, and global gene expression profiles. Although diversity has often been observed in *P. aeruginosa* populations from the same CF respiratory specimen, we show here that such diversity can have long-term stability as demonstrated by the decade long co-existence of the three different sub-lineages.

From an ecological perspective, these findings indicate population diversification as a result of adaptive radiation processes and niche partitioning. Evolutionary theory and experimental evolution studies – in which microbial populations evolve under defined, artificial conditions in the laboratory - have clearly established a connection between heterogeneous environments (*i.e.* introduction of distinct and vacant niches, each with different selective conditions) and the evolution and maintenance of population diversity (27-29). Indeed, this explanation has been used to rationalize the observed population diversity in relation to chronic CF infections, but since most studies on *P. aeruginosa* population diversity rely on inspection of isolates from sputum or BAL samples which lack spatial resolution, conclusions about niche partitioning has been difficult. In this study we obtained parallel samples from different compartments in the infected individual, including material sampled from the paranasal sinuses, which enabled us to demonstrate a clear spatial distribution of the sub-lineages in different niches. We found cluster A to inhabit the paranasal sinuses, whereas we identified cluster B and C in samples from the lung environment (*i.e*. in sputum samples and in BAL samples). Because of the striking absence of parallel genomic evolution between cluster B and C, which clearly indicate that they populate niches with different selective pressures, we suggest that clusters B and C occupy distinct niches in the lungs of the patient. A distribution of the three sub-lineages in different niches would result in lineage-specific sets of genes being under positive selection, which was indeed observed (Figure 2C). Importantly, recent documentations of spatial distributions of microorganisms in CF lungs further support this model. For example, spatial heterogeneity of microbial communities in infected CF lungs has been demonstrated recently by culture-independent analysis of CF lung tissue samples (30), and microscopy analysis of explanted CF lungs has furthermore shown that mucoid and non-mucoid variants of *P. aeruginosa* are located in different niches such as the sputum in the conductive zone of the bronchi and the respiratory zones of the alveoli and respiratory bronchioles in the infected CF lung, respectively (31). Interestingly, cluster B and C are different with respect to the mucoid phenotype (isolates from cluster C are mucoid as a consequence of mutations in *mucA,* and isolates from cluster B are non-mucoid as a consequence of mutations in both *mucA* and *algT*), and it is possible that a related distribution is present in the patient studied here. The observation that the faster growing cluster B isolates consistently

constituted only a small fraction of the population sampled recently from the lung environment and was even absent from the population in some cases, could be related to difficulties in sampling the particular lung niche occupied by cluster B. This is in accordance with observations from CF autopsies and explanted lungs, which showed that the non-mucoid phenotype was actually phagocytosed by PMNs in contrast to the biofilm-growing mucoid phenotype (34). Alternatively, or complementary, it is possible that isolates from cluster C are better adapted than cluster B to the stressful and selective conditions in the lung environment. Indeed, cluster C isolates exhibited a larger panel of phenotypes often found in isolates from chronic infections, and had a larger number of pathoadaptive mutations than cluster B. The prevalence of pathoadaptive mutations has previously been shown to be an indicator of ecological success (4), and the observation that cluster C had more pathoadaptive mutations in total than cluster B, provide a possible explanation to the differences in their population sizes in lung samples.

Our phylogenomic analysis furthermore showed that cluster B diverged from the population soon after or coinciding with initial colonization, and that the DK1 population further diverged ten years later. By combining this phylogeny with the spatial location of each of the three clusters it is now possible to reconstruct the evolutionary history of DK1 in the patient studied here. We propose that the patient was first colonized in the paranasal sinuses with isolates belonging to cluster A. Following initial colonization, isolates from the infected sinuses migrated to the lungs in two distinct events (separated by a decade) and evolved into highly successful colonizers of the lungs (*i.e*. cluster B and C). Based on observations of extensive overlap in colony morphotypes between clonal isolates sampled from the paranasal sinuses and BAL samples, we have previously demonstrated migration from the sinuses to the lung environment (9). The genomic data presented here provide further evidence for this hypothesis, and highlight the dynamic nature of the infecting population as well as the significance of the paranasal sinuses as functional pathogen reservoirs.

Interestingly, the amount of evolutionary changes was different in the three clusters. Isolates from cluster B and C accumulated more pathoadaptive mutations and demonstrated more phenotypic alterations as a function of time than the sinus-associated population in cluster A. We speculate that the observed differences are related to mutation pathways activated by stress responses (*i.e.* induced when cells are maladapted to their environment) (32, 33), differences in *in vivo* growth rates, or due to oxygen radical damage as a result of the inflammatory response which differs in different niches of the respiratory tract. Although the sinus and lung environments share some physiological properties such as thick mucus that can function as sites for bacterial growth, there are also significant differences that may explain the different evolutionary trajectories associated with each compartment. For example, the inflammatory response is reduced in the paranasal sinuses and polymorphonuclear leukocytes, which produce oxygen

radicals (34) dominates in the lungs and sputum whereas they are rare in the paranasal sinuses where the inflammation is less pronounced due to high production of Immunoglobulin A (35). Further, antibiotic penetration and hence the achievement of therapeutic levels is hypothesized to be less efficient in the sinuses than in the lungs (35, 36).

We have previously shown a much more pronounced evolutionary response in sinus-associated *P. aeruginosa* populations in intermittently colonized patients than what we show here for cluster A (9). In addition, studies have shown extensive overlaps in evolved phenotypes such as colony morphology and global gene expression profiles between isolates from sinus and lung samples from the same chronically infected patients (9, 36), which was clearly not observed here. These observational differences suggest that generalizing conclusions concerning evolution of *P. aeruginosa* in the paranasal sinuses should be considered with caution, and that our understanding of the factors that influence the evolutionary trajectories of sinus-associated *P. aeruginosa* populations is incomplete.

Our results show that the environmental heterogeneity of the CF airways contribute to the generation and stability of intraclonal diversity, and it is possible that similar processes underlie the population diversity observed in other CF-associated pathogens (13). More broadly, our study demonstrates that *P. aeruginosa* isolated from chronic CF patients are representatives of an underlying complex and dynamic population structure comprised of distinct clonal variants that are distributed in different niches. Our results suggest that assuming that one isolate in sputum and BAL samples is representative of the entire population may lead to an inaccurate description of therapy-relevant phenotypes such as antibiotic resistance profiles, and that treatment strategies targeting the dominant population need not be effective in other areas of the airways, or against other sub-populations. A more complete understanding of the spatial heterogeneity and population diversity could therefore have positive implications for efficient therapy and disease outcome.

Materials and methods

Ethics Statement. No patient samples were taken specifically for this study. Isolation and storage of bacterial isolates was part of the established routine at the Department of Clinical Microbiology, Rigshospitalet. The bacterial isolates used in the study included isolates from the bacterial collection available at the Department of Clinical Microbiology, Rigshospitalet as well as bacteria isolated during Functional Endoscopic Sinus Surgery (FESS) on patient P30M0 in 2012. The use of bacterial isolates was approved by the local ethics committee, Region Hovedstaden. All samples were anonymized.

Bacterial strains. The investigated strains of the *P. aeruginosa* DK1 clone type consisted of 18 longitudinal isolates from patient P30M0 and four historical isolates from 1973 from patient P33F0, P50F0, P13F1, and

P43M1. Patient P30M0 was chosen based on the long clonal infection history with DK1 and the availability of stored isolates and new lung expectorates. Additional clinical strains from P30M0 were isolated from either sinus or bronchoalveolar lavage (BAL) samples from 2012. The samples from the paranasal sinuses were taken during Functional Endoscopic Sinus Surgery (FESS) as previously described (37). Isolation and identification of *P. aeruginosa* was done as previously described (38). The isolates were grown in Luria-Bertani (LB) medium and stored at -80°C with 10% glycerol. The *P. aeruginosa* isolates investigated in this study were confirmed to belong to the DK1 lineage by two independent genotyping methods: pulsed-field gel electrophoresis and SNP genotyping using AT biochips (Clondiag Chip Technologies, Germany) as described previously (39).

Genome sequencing. Genomic DNA was prepared from *P. aeruginosa* isolates using Wizard Genomic DNA Purification Kit (Promega). Genomes were sequenced to an coverage depth of 90 fold (range 17 to 207 fold) on either an Illumina GAIIx platform generating 75-bp single-end reads (6 isolates), or on a HiSeq2000 platform generating either 75-nt single-end reads (7 isolates) or 100-nt paired-end reads (9 isolates). The genome of DK1-P33F0 was *de novo* assembled into 6,156,016 bp (252 contigs, N50=49,984 bp) with the assistance of *P. aeruginosa* genome sequences of PAO1, PA14, and LESB58, respectively, using the Columbus module of Velvet version 1.0.16 (40). The genome assembly of DK1-P33F0 was then used as a reference for alignment of sequence reads from each of the isolates to identify high-quality SNPs in the non-repetitive parts of the genome.

Mutation detection, analysis and construction of the phylogenetic tree were performed as previously described (4). Calculation of the d*N*/d*S* ratio was based on the assumption that a random mutation in the total number of possible SNP mutation would cause a synonymous change 25% of the time (41, 42). Bayesian analysis of evolutionary rates was performed using BEAST v1.7.2 (43) with a lognormal relaxed molecular clock model and a general time-reversible substitution model. Mutation rates were calculated from a chain length of 50 million steps, sampled every 5,000 steps. The first 5 million steps were discarded as a burn-in. A maximum clade credibility tree was generated using the TreeAnnotator program from the BEAST package), and the effective sample-sizes (ESS) of all parameters were >1,000 as calculated by Tracer v1.5 (available from http://beast.bio.ed.ac.uk/Tracer), which was also used to calculate 95% HPD confidence intervals of the divergence times (*i.e.* an interval in which the modeled parameter resides with 95% probability).

Biolog Phenotype MicroArray™. Phenotype MicroArray (Biolog, Hayward, CA) experiments were performed in duplicate according to manufacturer instructions (44, 45). Bacterial strains were streaked on LB agar plates and incubated at 37°C until colonies appeared on plates (23-48 h). Cells were swabbed from

the plates and suspended in IF-0 GN Base (inoculation fluid) at a density corresponding to 42% transmittance in the Biolog turbidimeter. The cell suspensions were diluted 1:6 in IF-0 minimal medium containing Biolog redox dye mixture D (tetrazolium), and 100 µL aliquots were added to carbon source plates (PM1). Nitrogen plates (PM3) were prepared with ferric citrate and glucose supplemented to the inoculation mixture. In total 190 substrates were available on the two plates. Of these, 125 were found to be informative by showing consistent growth between replicates. OmniLog *OL_FM/Kin* 1.20.02 software (Biolog) was used to export the OmniLog data and the average area beneath each kinetic curve was used for analysis. Total catabolic function was calculated as previously described(46). Further data and statistical analysis was performed using R version 2.10.1 (R Development Core Team 2009).

Affymetrix GeneChip. Strains were grown aerobically at 37°C in LB medium starting at OD_{600} = 0.01 and harvested at OD_{600} = 0.5. Immediately after harvesting, cells were mixed with either RNAprotect Bacteria Reagent (Qiagen) if non-mucoid cells, or mixed with STOP solution (95% ethanol, 5% phenol) if cells were mucoid to avoid clumping. After 10 minutes of centrifugation at 7000 g at 4°C the supernatant was discarded and cell pellet was stored at -80°C. RNA extraction, cDNA preparation and labelling, and hybridization were performed as previously described (42). The raw CEL files were obtained by the Affymetrix GeneChip Operating System 1.4 and the analysis was performed by using BioConductor tools in R environment (47). Strains were tested in triplicates.

Motility assays. The motility of all isolates was tested on ABT agar supplemented with 0.5% casamino acids and 0.5% glucose as previously described (48). Briefly, ABT plates were inoculated with single colonies on the top, in the middle or on the bottom of the plates to measure swarm, swim, or twitch, respectively. Swim and swarm were measured on 0.3% (wt/vol) agar and 0.6% (wt/vol), respectively and incubated 24 h at 30°C. Twitching motility was determined on 1.5% (wt/vol) agar and incubated 48 hours at 37°C. The motility zones were calculated in triplicates relative to the reference strain PAO1.

Detection of quorum sensing signals. For detection of QS signals *P. aeruginosa* strains were cross-streaked against two *E. coli* MH155 and MH205 monitor strains as previously described (48). After 48 h of incubation at 37°C the AHL production was detected by inspecting the fluorescence of the monitor strains with a Zeiss Axioplan 2 microscope with a 2.5/0.075× neofluar plan objective.

Skim milk protease assay. The production of protease was determined by applying one colony of the isolate grown on LB agar to skim milk plates (LB agar with 10% skim milk). The plates were incubated for 24 h at 37°C and the clearing zones were measured.

Determination of hypermutable phenotypes. The isolates were grown overnight in LB broth at 37°C. The hypermutability of the isolates was determined by plating aliquots of serial dilutions of bacteria cells on LB plates with and without Rifampicin (300 µg/mL) or Streptomycin (500 µg/mL). The plates were incubated at

37°C for 24 or 48 hours and the number of colonies was counted. An isolate was considered hypermutable if it had 20 times higher mutation frequency after exposure to rifampicin and streptomycin than the reference strain.

Growth rate measurement and estimation of *in situ* bacterial generations. LB broth was inoculated with the isolate and incubated at 37°C overnight. A cell culture volume was transferred into 50 mL LB in a 250 mL shaking flask with baffles generating an OD_{600nm} of 0.01. The flask was incubated at 37°C with shaking at 240 rpm in a New Brunswick Scientific incubator. Growth rates were measured by monitoring the OD_{600nm} during growth. The measurements were done at least in duplicates.

Growth rates were calculated from the exponential part of the growth curves and expressed as doubling times in minutes (ln(2)/minutes).

PCR screening. A multiplex PCR strategy was constructed for determination of different sub-lineages of *P. aeruginosa* from patient P30M0. Sub-lineage-specific primers were designed including an internal positive control for all lineages (Table S3). PCR amplifications were performed with the following optimized parameters: pre-denaturation for 5 min at 95°C; 25 cycles of 95°C for 30 s, 66°C for 30 s, and 72°C for 25 s; post-extension for 3 min at 72°C; and soaking at 4°C. PCR products were resolved in a 2% agarose gel in 1XTBE buffer at 100 V and visualized with ethidium bromide. In total 3*64 isolates were tested from the sinus, BAL, and sputum sample from 2012. Furthermore, isolates sampled from sputum samples from 2010, 2011, and 2012 were tested.

Acknowledgement

We would like to thank S. Damkiær for assistance with Affymetrix Genechips and Biolog, and N. Jochumsen for support with the multiplex PCR. We would also like to thank L. Yang (Department of Biological Engineering, Massachusetts Institute of Technology) and the Infection Microbiology Group at DTU for helpful discussions. U. Johansen is thanked for technical assistance.

This work was supported by the Danish Council for Independent Research and the Lundbeck Foundation. HKJ was supported by a clinical research stipend from the Novo Nordisk Foundation. LJ was supported by a Young Investigator grant from the Villum Foundation.

Author contributions: T.M, S.M, and L.J. conceived the study and designed research; T.M, R.L.M., M.G-L., and A.E.B. performed research; K.A., H.K.J and N.H. contributed materials; T.M., R.L.M., S.M., and L.J. analysed the data; T.M. and L.J. wrote the manuscript.

References

1. Høiby N (2011) Recent advances in the treatment of Pseudomonas aeruginosa infections in cystic fibrosis. *BMC Med* 9:32.
2. Jelsbak L*, et al.* (2007) Molecular Epidemiology and Dynamics of Pseudomonas aeruginosa Populations in Lungs of Cystic Fibrosis Patients. *Infect Immun* 75(5):2214-2224.
3. Nguyen D & Singh PK (2006) Evolving stealth: genetic adaptation of Pseudomonas aeruginosa during cystic fibrosis infections. *Proc Natl Acad Sci U S A* 103(22):8305-8306.
4. Marvig RL, Johansen HK, Molin S, & Jelsbak L (2013) Genome Analysis of a Transmissible Lineage of Pseudomonas aeruginosa Reveals Pathoadaptive Mutations and Distinct Evolutionary Paths of Hypermutators. *PLoS Genet* 9(9):e1003741.
5. Smith EE*, et al.* (2006) Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A* 103(22):8487-8492.
6. Cramer N*, et al.* (2011) Microevolution of the major common Pseudomonas aeruginosa clones C and PA14 in cystic fibrosis lungs. *Environ Microbiol* 13(7):1690-1704.
7. Wilder CN, Allada G, & Schuster M (2009) Instantaneous within-patient diversity of Pseudomonas aeruginosa quorum-sensing populations from cystic fibrosis lung infections. *Infect Immun* 77(12):5631-5639.
8. Rau MH*, et al.* (2010) Early adaptive developments of Pseudomonas aeruginosa after the transition from life in the environment to persistent colonization in the airways of human cystic fibrosis hosts. *Environ Microbiol* 12(6):1643-1658.
9. Hansen SK*, et al.* (2012) Evolution and diversification of Pseudomonas aeruginosa in the paranasal sinuses of cystic fibrosis children have implications for chronic lung infection. *ISME J* 6(1):31-45.
10. Mowat E*, et al.* (2011) Pseudomonas aeruginosa population diversity and turnover in cystic fibrosis chronic infections. *Am J Respir Crit Care Med* 183(12):1674-1679.
11. Workentine ML*, et al.* (2013) Phenotypic heterogeneity of Pseudomonas aeruginosa populations in a cystic fibrosis patient. *PLoS One* 8(4):e60225.
12. Chung JC*, et al.* (2012) Genomic variation among contemporary Pseudomonas aeruginosa isolates from chronically infected cystic fibrosis patients. *J Bacteriol* 194(18):4857-4866.
13. Lieberman TD*, et al.* (2014) Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nature genetics* 46(1):82-87.
14. Qin X*, et al.* (2012) Pseudomonas aeruginosa syntrophy in chronically colonized airways of cystic fibrosis patients. *Antimicrob Agents Chemother* 56(11):5971-5981.
15. Ciofu O, Mandsberg LF, Bjarnsholt T, Wassermann T, & Høiby N (2010) Genetic adaptation of Pseudomonas aeruginosa during chronic lung infection of patients with cystic fibrosis: strong and weak mutators with heterogeneous genetic backgrounds emerge in mucA and/or lasR mutants. *Microbiology* 156(Pt 4):1108-1119.
16. Jelsbak L*, et al.* (2007) Molecular epidemiology and dynamics of Pseudomonas aeruginosa populations in lungs of cystic fibrosis patients. *Infect Immun* 75(5):2214-2224.
17. Oliver A, Canton R, Campo P, Baquero F, & Blazquez J (2000) High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. *Science* 288(5469):1251-1254.
18. Pedersen SS, Høiby N, Espersen F, & Koch C (1992) Role of alginate in infection with mucoid Pseudomonas aeruginosa in cystic fibrosis. *Thorax* 47(1):6-13.
19. Martin DW*, et al.* (1993) Mechanism of conversion to mucoidy in Pseudomonas aeruginosa infecting cystic fibrosis patients. *Proc Natl Acad Sci U S A* 90(18):8377-8381.
20. Yang L*, et al.* (2008) In situ growth rates and biofilm development of Pseudomonas aeruginosa populations in chronic lung infections. *J Bacteriol* 190(8):2767-2776.
21. Hoffman LR*, et al.* (2009) Pseudomonas aeruginosa lasR mutants are associated with cystic fibrosis lung disease progression. *J Cyst Fibros* 8(1):66-70.

22. Mahenthiralingam E, Campbell ME, & Speert DP (1994) Nonmotility and phagocytic resistance of Pseudomonas aeruginosa isolates from chronically colonized patients with cystic fibrosis. *Infect Immun* 62(2):596-605.
23. Thogersen JC, Morup M, Damkiaer S, Molin S, & Jelsbak L (2013) Archetypal analysis of diverse Pseudomonas aeruginosa transcriptomes reveals adaptation in cystic fibrosis airways. *BMC Bioinformatics* 14(1):279.
24. Huse HK*, et al.* (2010) Parallel Evolution in Pseudomonas aeruginosa over 39,000 Generations In Vivo. *MBio* 1(4).
25. Bielecki P*, et al.* (2012) Mutation in Elongation Factor G Confers Resistance to the Antibiotic Argyrin in the Opportunistic Pathogen Pseudomonas aeruginosa. *Chembiochem : a European journal of chemical biology*.
26. Pasca MR*, et al.* (2012) Evaluation of fluoroquinolone resistance mechanisms in Pseudomonas aeruginosa multidrug resistance clinical isolates. *Microb Drug Resist* 18(1):23-32.
27. Rainey PB & Travisano M (1998) Adaptive radiation in a heterogeneous environment. *Nature* 394(6688):69-72.
28. Hansen SK, Rainey PB, Haagensen JA, & Molin S (2007) Evolution of species interactions in a biofilm community. *Nature* 445(7127):533-536.
29. Boles BR, Thoendel M, & Singh PK (2004) Self-generated diversity produces "insurance effects" in biofilm communities. *Proc Natl Acad Sci U S A* 101(47):16630-16635.
30. Willner D*, et al.* (2012) Spatial distribution of microbial communities in the cystic fibrosis lung. *ISME J* 6(2):471-474.
31. Bjarnsholt T*, et al.* (2009) Pseudomonas aeruginosa biofilms in the respiratory tract of cystic fibrosis patients. *Pediatr Pulmonol* 44(6):547-558.
32. Boles BR & Singh PK (2008) Endogenous oxidative stress produces diversity and adaptability in biofilm communities. *Proc Natl Acad Sci U S A* 105(34):12503-12508.
33. Bjedov I*, et al.* (2003) Stress-induced mutagenesis in bacteria. *Science* 300(5624):1404-1409.
34. Kolpen M*, et al.* (2010) Polymorphonuclear leucocytes consume oxygen in sputum from chronic Pseudomonas aeruginosa pneumonia in cystic fibrosis. *Thorax* 65(1):57-62.
35. Johansen HK*, et al.* (2012) Colonisation and infection of the paranasal sinuses in cystic fibrosis patients is accompanied by a reduced PMN response. *J Cyst Fibros* 11(6):525-531.
36. Ciofu O*, et al.* (2013) P. aeruginosa in the paranasal sinuses and transplanted lungs have similar adaptive mutations as isolates from chronically infected CF lungs. *J Cyst Fibros* 12(6):729-736.
37. Aanaes K*, et al.* (2013) Clinical effects of sinus surgery and adjuvant therapy in cystic fibrosis patients - can chronic lung infections be postponed? *Rhinology* 51(3):222-230.
38. Høiby N & Frederiksen B (2000) Microbiology of cystic fibrosis. *Cystic fibrosis*, (Arnold, London United Kingdom), pp 83-107.
39. Wiehlmann L*, et al.* (2007) Population structure of Pseudomonas aeruginosa. *Proc Natl Acad Sci U S A* 104(19):8101-8106.
40. Zerbino DR & Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* 18(5):821-829.
41. Yang Z & Bielawski JP (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* 15(12):496-503.
42. Yang L*, et al.* (2011) Evolutionary dynamics of bacteria in a human host environment. *Proc Natl Acad Sci U S A* 108(18):7481-7486.
43. Drummond AJ, Suchard MA, Xie D, & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969-1973.
44. Bochner BR (2003) New technologies to assess genotype-phenotype relationships. *Nature reviews. Genetics* 4(4):309-314.
45. Bochner BR (2009) Global phenotypic characterization of bacteria. *FEMS microbiology reviews* 33(1):191-205.

46. Cooper VS & Lenski RE (2000) The population genetics of ecological specialization in evolving Escherichia coli populations. *Nature* 407(6805):736-739.
47. Gentleman RC*, et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5(10):R80.
48. Hentzer M*, et al.* (2002) Inhibition of quorum sensing in Pseudomonas aeruginosa biofilm bacteria by a halogenated furanone compound. *Microbiology* 148(Pt 1):87-102.

Figures:

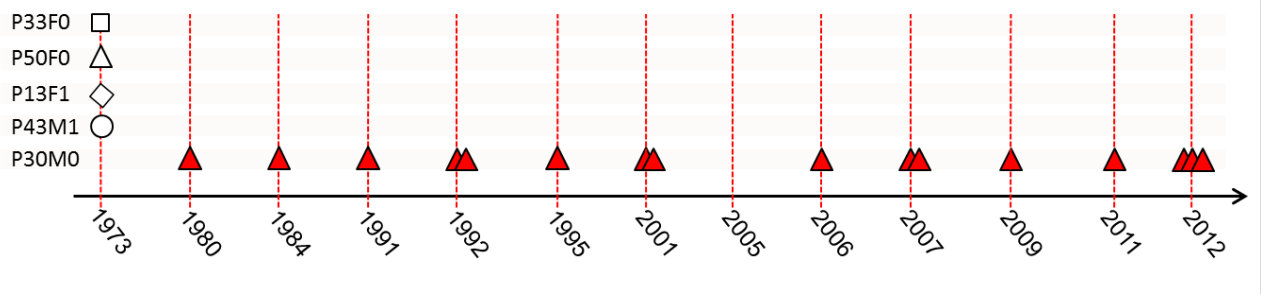
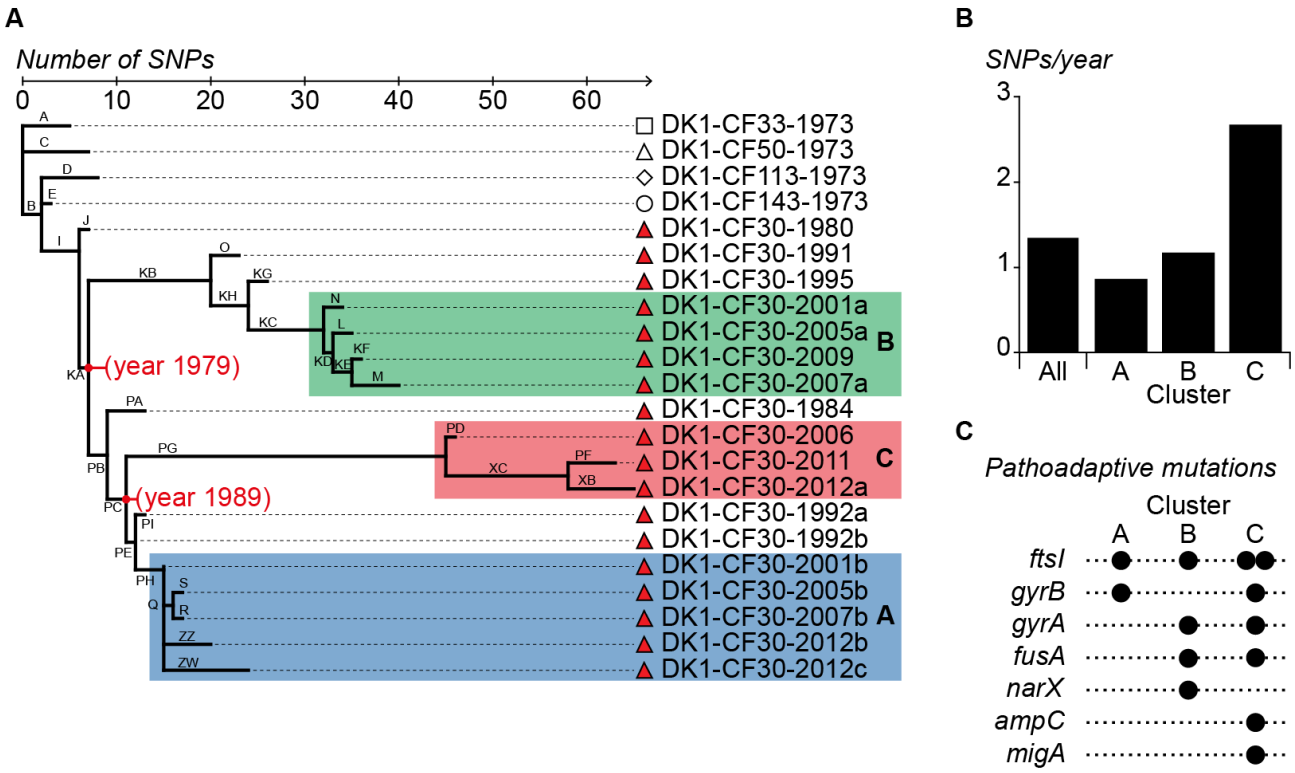


Figure 1. Patient origin and sampling time of genome sequenced *P. aeruginosa* DK1 isolates. DK1 *P. aeruginosa* isolates were sampled from five different CF patients between 1973 and 2012. Bacterial isolates from different patients are specified by symbols. Stacked symbols indicate that multiple isolates were sampled the same year from the same patients. The isolates are named by their clone type, and by the patient from whom they were isolated and their isolation year (e.g. DK1-P30M0-1980).



549

550 **Figure 2. Evolutionary trajectory of the DK1 lineage.** A) Maximum-parsimonious reconstruction of the
551 phylogeny of DK1 clones isolated from CF patients P30M0, P33F0, P50F0, P13F1, and P43M1. The tree is
552 based on 160 SNPs (identified from genome sequencing) which accumulated in a highly parsimonious
553 fashion (parsimonious consistency of 0.99). Capital letters indicate branch names, and lengths of branches
554 are proportional to the number of SNPs. The specific mutations that have accumulated during each specific
555 branch are listed in Table S1. The three phylogenetic clusters in P30M0 are labelled A, B, and C. Median
556 estimates of divergence dates (in calendar years) are given in red for the major nodes. B) Mutation rates
557 (SNPs per year) for each of the three clusters and for all DK1 isolates. C) Prevalence of pathoadaptive
558 mutations in the three DK1 sub-lineages. The genes hit by nonsynonymous SNPs are indicated with filled
559 circles for each of the three DK1 sub-lineages in P30M0.

560

A

Isolate/year	Morphology[a]	Motility[b]		QS[c]		Protease[d]
		Twitch	Swarm	3-O-C_{12}-HSL	C_4-HSL	
PAO1	n	++++	++++	++	++	+++
DK1-P33F0-1973	n	+++	-	++	++	++
DK1-P50F0-1973	n	++	-	++	++	++
DK1-P13F1-1973	n	++	-	++	++	++
DK1-P43M1-1973	n	++	-	++	++	++
DK1-P30M0-1980	n	-	-	++	++	++
DK1-P30M0-1984	n	++	-	++	++	++
DK1-P30M0-1991	n	-	-	++	++	++
DK1-P30M0-1992a	n	-	-	++	++	++
DK1-P30M0-1995	n	-	-	+	++	+
DK1-P30M0-2001a	n	+	-	++	++	+
DK1-P30M0-2005a	n	-	-	++	++	+
DK1-P30M0-2007a	n	-	-	++	++	+
DK1-P30M0-2009	n	-	-	++	++	+
DK1-P30M0-1992b	m	-	+	++	++	+
DK1-P30M0-2001b	m	-	-	++	++	-
DK1-P30M0-2005b	m	-	-	++	++	-
DK1-P30M0-2006	m	-	-	-	-	-
DK1-P30M0-2007b	m	-	-	++	++	-
DK1-P30M0-2011	m	-	-	-	-	-

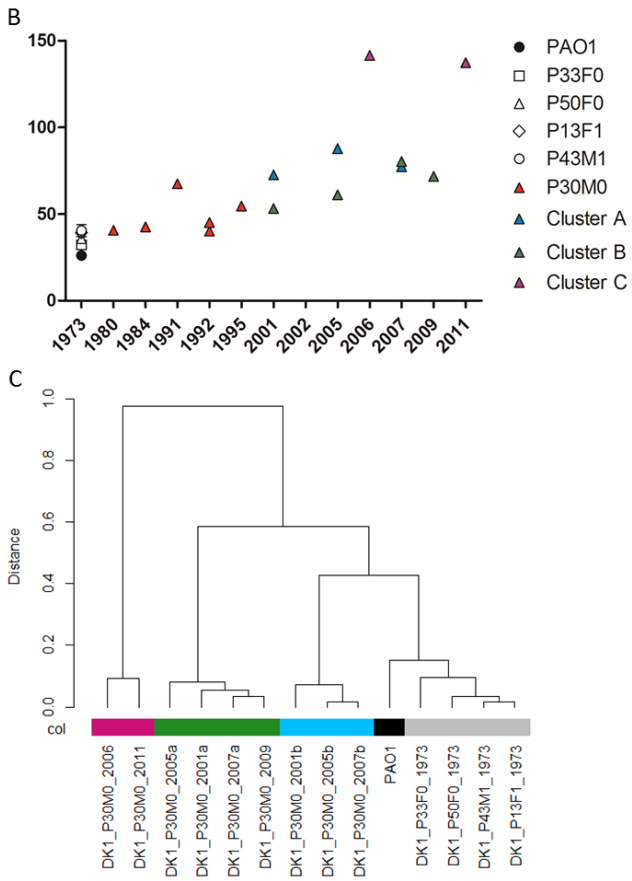
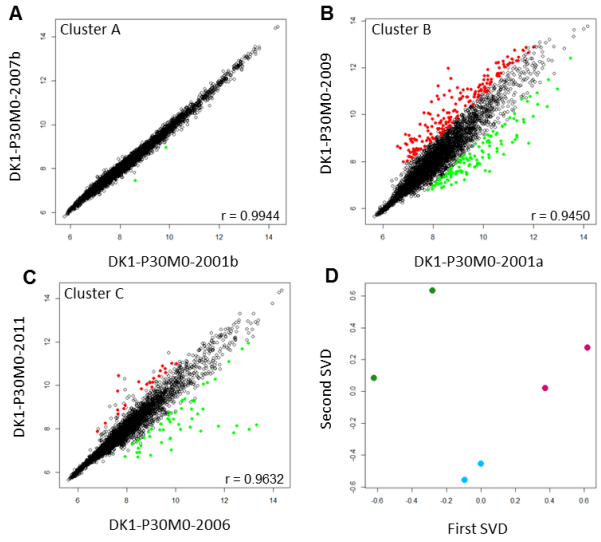


Figure 3. Phenotypic characterization of DK1 isolates. A) [a]nm, non-mucoid; m, mucoid. [b]Motility relative to PAO1. [c]Production of AHLs detected by inspection of bioluminescence of monitor strain; ++, high level; +, low level; -, not detectable. [d]Detection of proteases on skim milk plates; +++, high level; +, low level; -, not detectable. B) Doubling time of historical isolates and P30M0 isolates measured in LB medium. C) Dendrogram showing the hierarchical cluster analysis (Wards linkage, Euclidean Distance) of global catabolic function. Colours represent each cluster; cluster A, blue; cluster B, green; and cluster C, purple.



569

570 **Figure 4. Transcriptome analysis.** A) Comparison of gene expression of the two mucoid isolates DK1-
571 P30M0-2001b and DK1-P30M0-2007b from cluster A. B) Comparisons of the two non-mucoid isolates DK1-
572 P30M0-2001a and DK1-P30M0-2009a from cluster B. C) Comparison of gene expression of the two mucoid
573 isolates DK1-P30M0-2006 and DK1-P30M0-2011 from cluster C. The red dots represent genes with
574 significantly increased expression (fold change ≥ 2, $P \leq 0.05$), and green dots indicate genes with decreased
575 expression. The r (Pearson correlation) values indicate the strength and direction of linear relationship
576 between the expression levels of the isolates. D) Single value decomposition showing the gene expression
577 relationships among isolates from cluster A, B and C. Each circle represents the mean of duplicates. Blue
578 circles represent isolates from cluster A, green circles represent isolates from cluster B, and isolates from
579 cluster C are shown as purple circles.

580

Table S1: SNP changes observed in the P30M0 lineage.

Branch	Position	Type of mutation	Gene name		Product	PseudoCAP Function Class	Independt mutations observed within the gene
listA	T163P	Missense	PA0652	vfr	transcriptional regulator Vfr	Transcriptional regulators	1
listA	L7L	Silent	PA4695	ilvH			1
listA	R12L	Missense	PA1430	lasR	transcriptional regulator LasR		1
listA	G476S	Missense	PA0928	gacS	sensor/response regulator hybrid	Two-component regulatory systems	1
listA	A19P	Missense	PA2020		probable transcriptional regulator	Transcriptional regulators	1
listB	G505A	Missense	PA3545	algG	alginate-c5-mannuronan-epimerase AlgG		2
listB	G10G	Silent	PA4755	greA	transcription elongation factor GreA	Transcription, RNA processing and degradation	1
listC	P515P	Silent	PA14_03220		Hypothetical protein	Hypothetical, unclassified, unknown	2
listC	S480S	Silent	PA0262		conserved hypothetical protein	Hypothetical, unclassified, unknown	1
listC	31	Not annotated					0
listC	A28P	Missense	PA3897		hypothetical protein	Membrane proteins	1
listC	Y59C	Missense	PA0762	algU	sigma factor AlgU	Transcriptional regulators	5
listC	R527H	Missense	PA3257	prc	periplasmic tail-specific protease		1
listC	T57T	Silent	PA3399		hypothetical protein	Hypothetical, unclassified, unknown	1
listD	G73G	Silent	PA4286		hypothetical protein	Hypothetical, unclassified, unknown	1
listD	A380A	Silent	PA1238				1
listD	K55K	Silent	PA3127		hypothetical protein	Hypothetical, unclassified, unknown	1
listD	A168V	Missense	PA0021		conserved hypothetical protein	Membrane proteins	1
listD	L10Q	Missense	PA5537		hypothetical protein	Hypothetical, unclassified, unknown	1
listD	L435L	Silent	PA2840		probable ATP-dependent RNA helicase	Transcription, RNA processing and degradation	1
listE	I65I	Silent	PA1192		conserved hypothetical protein	Hypothetical, unclassified, unknown	1
listI	L117F	Missense	PA2196		probable transcriptional regulator	Transcriptional regulators	1
listI	11717	Intergenic	PA1566/I	/	15 upstream conserved hypothetical protein/453	Hypothetical, unclassified, unknown/Hypothetic	1
listI	N208N	Silent	PA1588	sucC	succinyl-CoA synthetase beta chain	Energy metabolism	1
listI	D184N	Missense	PA1775	cmpX		Membrane proteins	2
listJ	D18G	Missense	PA0762	algU	sigma factor AlgU	Transcriptional regulators	5
listKA	R504C	Missense	PA4418	ftsI	penicillin-binding protein 3	Cell division; Cell wall / LPS / capsule	2
listKB	G344G	Silent	PA4900				1
listKB	T122T	Silent	PA4089		probable short-chain dehydrogenase	Putative enzymes	1
listKB	235	Silent	PA1094	fliD	flagellar capping protein FliD		1
listKB	V193M	Missense	PA2329			Transport of small molecules	1
listKB	V849G	Missense	PA3168	gyrA	DNA gyrase subunit A		2
listKB	G118S	Missense	PA4266	fusA1	elongation factor G		2
listKB	G301R	Missense	PA4860		probable permease of ABC transporter		1
listKB	A379V	Missense	PA0835	pta	phosphate acetyltransferase	Carbon compound catabolism	1
listKB	H162R	Missense	PA4954	motA	chemotaxis protein MotA	Adaptation, Protection; Chemotaxis	1
listKB	T213T	Silent	PA1818	ldcA		Amino acid biosynthesis and metabolism	1
listKB	E182K	Missense	PA1810			Transport of small molecules	1
listKB	P230P	Silent	PA4811	fdnH		Energy metabolism	1
listKB	R243C	Missense	PA4843		probable two-component response regulator		1
listKC	A8T	Missense	PA3951		conserved hypothetical protein	Hypothetical, unclassified, unknown	1
listKC	H261L	Missense	PA3878	narX	two-component sensor NarX		1
listKC	D292A	Missense	PA5320	coaC			2
listKC	L39L	Silent	PA0761	nadB	L-aspartate oxidase		1
listKC	L32L	Silent	PA1121	yfiR	YfiR	Cell wall / LPS / capsule	1
listKC	Y196_	Nonsense	PALES_4	oprD		Transport of small molecules	1
listKC	F190F	Silent	PA2069		probable carbamoyl transferase	Putative enzymes	1
listKC	P36L	Missense	PA2674		probable type II secretion system protein	Protein secretion/export apparatus	1
listKD	124819	Intergenic	PA2138/I	/	16 downstream probable ATP-dependent DNA lig	/Hypothetical, unclassified, unknown	1
listKE	P353L	Missense	PA3162	rpsA	30S ribosomal protein S1		1
listKE	8641	Intergenic	PA1563/I	/	38 downstream conserved hypothetical protein/4	Hypothetical, unclassified, unknown/Hypothetic	1
listKF	R160R	Silent	PA2156		conserved hypothetical protein	Hypothetical, unclassified, unknown	1
listKG	4965	Not annotated					0
listKG	20159	Intergenic	PA0915/I	/	59 upstream conserved hypothetical protein/3 do	Hypothetical, unclassified, unknown/Hypothetic	1
listKH	T84P	Missense	PA4310	pctB	chemotactic transducer PctB	Adaptation, Protection; Chemotaxis	1
listKH	7743	Intergenic	PA4610/I	/	45 downstream hypothetical protein/17 downstre	Hypothetical, unclassified, unknown/Hypothetic	1
listKH	L666L	Silent	PA1181		conserved hypothetical protein	Membrane proteins	1
listKH	V255V	Silent	PA0687		probable type II secretion system protein	Protein secretion/export apparatus	1
listL	E572K	Missense	PA14_03220		Hypothetical protein	Hypothetical, unclassified, unknown	2
listL	R40C	Missense	PA4499		probable transcriptional regulator	Transcriptional regulators	1
listM	R21P	Missense	PA2537		probable acyltransferase	Fatty acid and phospholipid metabolism	1
listM	G107R	Missense	PA1923		hypothetical protein		1
listM	S190S	Silent	PA4809	fdhE	FdhE protein	Energy metabolism	1
listM	T345P	Missense	PA3464		hypothetical protein	Hypothetical, unclassified, unknown	1
listM	N545K	Missense	PA1992	ercS	ErcS	Two-component regulatory systems	1
listN	G641G	Silent	PA4021		probable transcriptional regulator	Transcriptional regulators	1
listN	Y258Y	Silent	PA0091	vgrG1	VgrG1	Hypothetical, unclassified, unknown	1
listO	A431T	Missense	PA5338	spoT			1
listO	Y59_	Nonsense	PA0762	algU	sigma factor AlgU	Transcriptional regulators	5
listO	G72G	Silent	PA0136			Transport of small molecules	1
listPA	P113L	Missense	PA2687	pfeS	two-component sensor PfeS	Two-component regulatory systems	1
listPA	A3021A	Silent	PA2424	pvdL	PvdL	Adaptation, Protection	2
listPA	S173L	Missense	PA0762	algU	sigma factor AlgU	Transcriptional regulators	5
listPA	R40Q	Missense	PA3304		conserved hypothetical protein	Hypothetical, unclassified, unknown	1
listPB	A1060A	Silent	PA14_36730		putative trehalose synthase	Carbon compound catabolism	1
listPB	L49L	Silent	PA4815		hypothetical protein	Membrane proteins	1
listPC	L273L	Silent	PA2776		conserved hypothetical protein	Hypothetical, unclassified, unknown	1
listPC	A115A	Silent	PA2925	hisM	histidine transport system permease HisM		1
listPD	G345D	Missense	PA2597		hypothetical protein	Putative enzymes	1
listPE	G308S	Missense	PA0368		conserved hypothetical protein	Hypothetical, unclassified, unknown	1
listPF	1786	Intergenic	PA4109/I	ampR/	104 upstream transcriptional regulator AmpR/45	Transcriptional regulators/Adaptation, Protectio	1
listPF	A55V	Missense	PA0607	rpe	ribulose-phosphate 3-epimerase	Energy metabolism	1
listPF	A775A	Silent	PA0903	alaS	alanyl-tRNA synthetase		1
listPF	29218	Intergenic	PA2330/I	/	7 upstream hypothetical protein/43 downstream	Hypothetical, unclassified, unknown/Membrane	1
listPF	I208I	Silent	PA5320	coaC			2
listPG	A289V	Missense	PA2998	nqrB		Energy metabolism	1
listPG	D277N	Missense	PA3547	algL	poly(beta-d-mannuronate) lyase precursor AlgL		1

listPG	V278M	Missense	PA4102	bfmS	BfmS		1
listPG	11767	Intergenic	PA1499/I/		256 upstream conserved hypothetical protein/41	Putative enzymes/Putative enzymes	1
listPG	Q185_	Nonsense	PA4410	ddlB	D-alanine--D-alanine ligase	Cell wall / LPS / capsule	1
listPG	G257R	Missense	PA4418	ftsI	penicillin-binding protein 3	Cell division; Cell wall / LPS / capsule	2
listPG	G75S	Missense	PA3168	gyrA	DNA gyrase subunit A		2
listPG	17777	Intergenic	PA3168/I	gyrA/	228 upstream DNA gyrase subunit A/9 downstrea /		1
listPG	K95N	Missense	PA4248	rplF	50S ribosomal protein L6		1
listPG	D467G	Missense	PA4266	fusA1	elongation factor G		2
listPG	G132G	Silent	PA5103		hypothetical protein	Hypothetical, unclassified, unknown	1
listPG	F779S	Missense	PA0004	gyrB	DNA gyrase subunit B		2
listPG	85377	Not annotated					0
listPG	H104Y	Missense	PA1167		hypothetical protein	Hypothetical, unclassified, unknown	1
listPG	Q402_	Nonsense	PA0411	pilJ	twitching motility protein PilJ	Chemotaxis; Motility & Attachment	1
listPG	P221P	Silent	PA1459		probable methyltransferase	Chemotaxis	1
listPG	N141D	Missense	PA1775	cmpX		Membrane proteins	2
listPG	T1282I	Missense	PA2424	pvdL	PvdL	Adaptation, Protection	2
listPG	S59L	Missense	PA2426	pvdS	sigma factor PvdS	Transcriptional regulators	1
listPG	S280L	Missense	PA3267		hypothetical protein	Membrane proteins	1
listPG	P71L	Missense	PA3385	amrZ	alginate and motility regulator Z		1
listPG	P358P	Silent	PA3339	PlpD	patatin-like protein, PlpD		1
listPG	A28A	Silent	PA5039	aroK	shikimate kinase	Amino acid biosynthesis and metabolism	1
listPG	P149L	Missense	PA5034	hemE	uroporphyrinogen decarboxylase		1
listPG	G543E	Missense	PA3141	wbpM	nucleotide sugar epimerase/dehydratase WbpM		1
listPG	4892	Not annotated					0
listPG	R61H	Missense	PA0337	ptsP		Transport of small molecules	2
listPG	A72T	Missense	PA0204		probable permease of ABC transporter		1
listPG	A419A	Silent	PA2652		probable chemotaxis transducer	Adaptation, Protection; Chemotaxis	1
listPG	V135M	Missense	PA2620	clpA	ATP-binding protease component ClpA		1
listPG	R84H	Missense	PA2837		probable outer membrane protein precursor	Transport of small molecules	1
listPG	E267K	Missense	PA5241	ppx	exopolyphosphatase		1
listPG	R19R	Silent	PA4785		probable acyl-CoA thiolase	Putative enzymes	1
listPG	P330P	Silent	PA1255		hypothetical protein	Hypothetical, unclassified, unknown	1
listPH	T296T	Silent	PA5013	ilvE	branched-chain amino acid transferase	Amino acid biosynthesis and metabolism	1
listPH	S466A	Missense	PA0004	gyrB	DNA gyrase subunit B		2
listPH	L345L	Silent	PA2059		probable permease of ABC transporter		1
listPI	D18G	Missense	PA0762	algU	sigma factor AlgU	Transcriptional regulators	5
listQ	25053	Not annotated					0
listR	E258Q	Missense	PA4670	prs	ribose-phosphate pyrophosphokinase		1
listS	M332I	Missense	PA5035	gltD	glutamate synthase small chain	Amino acid biosynthesis and metabolism	1
listXB	A112T	Missense	PA5133		conserved hypothetical protein	Membrane proteins	1
listXB	22058	Intergenic	PA1578/I/		10 downstream hypothetical protein/482 upstrea	Hypothetical, unclassified, unknown/Hypothetic	1
listXB	V163G	Missense	PA1833		probable oxidoreductase	Putative enzymes	1
listXB	S36_	Nonsense	PA4661	pagL	Lipid A 3-O-deacylase	Hypothetical, unclassified, unknown	1
listXB	H269Q	Missense	PA4666	hemA	glutamyl-tRNA reductase		1
listXB	D106N	Missense	PA0705	migA	alpha-1,6-rhamnosyltransferase MigA	Putative enzymes; Cell wall / LPS / capsule	1
listXB	72043	Intergenic	PA2894/I/		40 upstream hypothetical protein/306 downstrea	Hypothetical, unclassified, unknown/Hypothetic	2
listXC	H215Y	Missense	PA4110	ampC	beta-lactamase precursor	Adaptation, Protection	1
listXC	P534S	Missense	PA3545	algG	alginate-c5-mannuronan-epimerase AlgG		2
listXC	H353H	Silent	PA1331		conserved hypothetical protein	Hypothetical, unclassified, unknown	1
listXC	Q37Q	Silent	PA3103	xcpR	general secretion pathway protein E	Protein secretion/export apparatus	1
listXC	A461V	Missense	PA4228	pchD	pyochelin biosynthesis protein PchD		1
listXC	19417	Intergenic	PA4709/I/	phuR	128 upstream probable hemin degrading factor/5	Putative enzymes/Transport of small molecules	1
listXC	11844	Intergenic	PA5136/I/		12 downstream hypothetical protein/25 downstre	Hypothetical, unclassified, unknown/Hypothetic	1
listXC	V452V	Silent	PA5554	atpD	ATP synthase beta chain	Energy metabolism	1
listXC	R25H	Missense	PA2109		hypothetical protein	Hypothetical, unclassified, unknown	1
listXC	R808H	Missense	PA3297		probable ATP-dependent helicase	Transcription, RNA processing and degradation	1
listXC	G94D	Missense	PA3432		hypothetical protein	Membrane proteins	1
listXC	I16I	Silent	PA4827		arylamine N-acetyltransferase	Adaptation, Protection; Putative enzymes	1
listXC	T250M	Missense	PA3593		probable acyl-CoA dehydrogenase	Putative enzymes	1
listZZ	A71A	Silent	PA1095		hypothetical protein	Hypothetical, unclassified, unknown	1
listZZ	P31P	Silent	PA14_58740		hypothetical protein	Hypothetical, unclassified, unknown	1
listZZ	I61I	Silent	PA2185	katN	non-heme catalase KatN	Adaptation, Protection	1
listZZ	T74I	Missense	PA1583	sdhA	succinate dehydrogenase (A subunit)	Energy metabolism	1
listZZ	S372S	Silent	PA0172		hypothetical protein	Hypothetical, unclassified, unknown	1
listZW	A182A	Silent	PA3541	alg8	alginate biosynthesis protein Alg8		1
listZW	Q10_	Nonsense	PA0789		probable amino acid permease		1
listZW	H102Q	Missense	PA4726	cbrB	two-component response regulator CbrB		1
listZW	V79V	Silent	PA0749		hypothetical protein	Hypothetical, unclassified, unknown	1
listZW	A192V	Missense	PA3232		probable nuclease	Putative enzymes	1
listZW	T149P	Missense	PA1696	pscO	translocation protein in type III secretion	Protein secretion/export apparatus	1
listZW	L42L	Silent	PA0337	ptsP		Transport of small molecules	2
listZW	56672	Intergenic	PA2634/I	aceA/	178 downstream isocitrate lyase AceA/1 upstrean	Putative enzymes/Hypothetical, unclassified, unl	1
listZW	72043	Intergenic	PA2894/I/		40 upstream hypothetical protein/306 downstrea	Hypothetical, unclassified, unknown/Hypothetic	2

Branch names are denoted in the phylogenetic tree in Figure 2.

Evolutionary modification of *Pseudomonas aeruginosa* Vfr functionality during chronic infection

Trine M. Markussen[1], Tine Nielsen, Søren Molin[1*] and Lars Jelsbak[1*]

[1]Department of Systems Biology, Technical University of Denmark, DK-2800, Lyngby, Denmark

[2]Department of Clinical Microbiology, University Hospital, Rigshospitalet, 2100 Copenhagen

* Corresponding authors: Lars Jelsbak and Søren Molin, Department of Systems Biology, Technical University of Denmark, 2800 Lyngby, Denmark. Email: lj@bio.dtu.dk (Lars Jelsbak) and sm@bio.dtu.dk (Søren Molin). Telephone: +45 45256129.

Abstract

The phenotype of infecting environmental *P. aeruginosa* has been shown to change significantly during chronic infection. Loss or reduction of several virulence factors important for initiation of acute infections is selected for during long-term evolution to the CF airways. Mutations in the virulence factor regulator (*vfr*) gene have been consistently identified in clinical isolates from CF patients, indicating that *vfr* is a common mutational target in the adaptation of *P. aeruginosa* to the CF airways.

The adaptation and evolutionary dynamics of the dominant and highly successful *P. aeruginosa* DK1 clone has been investigated recently revealing early fixation of a mutation in the global transcriptional regulator Vfr (substitution).

Here we investigate the dominant *vfr* mutation that emerged in the DK1 clone during chronic infection by moving the mutation to the genome of the reference strain PAO1 and examine the phenotypic consequences. Our findings showed that a single naturally occurring mutation resulted in altered phenotypic expression of several virulence factors including protease secretion, production of pyocyanin, and twitching and swimming motility. It was demonstrated that the substitution did not result in a knockout Vfr but caused a functional modification in the Vfr protein that resulted in alternative regulation.

We show evidence of bacterial rewiring of the regulatory network by mutation in Vfr resulting in a phenotype resembling late chronic isolates, indicating that the role of Vfr in the regulation of virulence factors are quite complex and that mutations in *vfr* may confer selective advantages of *P. aeruginosa* in the CF environment during adaptation by fine-tuning of regulatory networks.

Introduction

Pseudomonas aeruginosa cause chronic airway infections in patients with cystic fibrosis (CF) and is the main cause of morbidity and mortality in these patients. *P. aeruginosa* acquires mutations and other genetic changes during the course of chronic infection in response to the selective forces in the CF airways. Hence the phenotypic characteristics of a chronic isolate sampled after long-term infection is significantly different from the initially infecting strain. While initially infecting strains often resemble environmental phenotypes e.g. fast growing, non-mucoid, and susceptible to antibiotics [1], chronic isolates display slow growth, mucoidy, antibiotic resistance, loss or reduction of multiple virulence factors, lack of motility, loss of quorum sensing and changed nutrition utilization [2-5].

It remains a central interest to understand the effects of such mutations on the physiology and behaviour of infecting bacterial isolates. It has been proposed that regulatory genes rather than structural genes play important roles during adaptation [6], as mutations in transcription regulation can shape entirely new traits by altering the function or activity of one central protein involved in a complex regulatory network and therefore have widespread pleiotropic effects.

We observed previously, that early regulatory mutations and their epistatic interactions indeed were important for the adaptation of the dominant *P. aeruginosa* DK2 clone contributing to the major phenotypic changes [7]. Experimental evolution studies have likewise documented that transcriptional regulators often are targets for adaptive mutations [8, 9].

Recent investigation detected an early and persistent mutation in the virulence factor regulator (*vfr*) gene observed in all DK1 isolates after 1973 [10]. Here we investigate this first regulatory mutation observed in the DK1 lineage.

The transcriptional regulator Vfr is a homolog of the cyclic AMP (cAMP) receptor protein (CRP) in *E. coli*, being 67 % identical and 91% similar to CRP [11]. The cloned *vfr* gene was able to complement a *crp* deficient *E. coli* mutant [11], suggesting functional similarities between these two proteins. Furthermore, amino acids important for CRP cAMP binding, DNA binding and bending, and RNA polymerase interaction were identical or conserved in Vfr [11]. Depending on the protein-protein interaction with RNA polymerase, CRP promoters can be divided into two classes [12]. At class I CRP-dependent promoters CRP bind upstream of the RNA polymerase and interact with α-CTD domain involving Activating Region 1 (AR1), which is a surface-exposed loop comprising residues 156 to 164 [13, 14]. At class II promoters CRP binds to a site overlapping the RNA polymerase binding site. The interaction involves both AR1 and α-CTD interaction but also AR2 (residues 19, 21, 96, and 101) interaction with the α subunit of the N-terminal of RNA polymerase [15]. In addition to these two classes there exists a third class of CRP promoters that require multiple CRP dimers and/or other activator molecules [12].

Vfr has been shown to control expression of over 200 genes [16, 17]. Vfr positively regulate the production of several virulence factors important for acute infection mainly through the regulation of the *las* quorum sensing (QS) system, as it binds directly to a DNA sequence within the promoter of *lasR* [18]. In most cases Vfr requires the binding of cAMP to be activated like the *E. coli* CRP protein, however, cAMP is not required for Vfr binding to the *lasR* promoter [19]. Vfr has also been found to bind directly within promoter regions of *toxA* (encoding exotoxin A), *regA*, *ptxR* (encoding *toxA* regulators), *fleQ* (encoding flagellar transcriptional regulator FleQ), and *cpdA* (encoding a cAMP phosphodiesterase) [20-23] and positively or negatively regulating the transcription of these genes. Furthermore, Vfr has been suggested to negatively regulate the PQS quorum sensing systems [24]. A recent study showed that Vfr regulates *rhlR* transcription by binding to several Vfr-binding sites present in the *rhlR* promoter region, one of which negatively regulates *rhlR* [25]. These findings suggest that Vfr is implicated in the complex network of regulating acute virulence factors and the quorum sensing system. The mutation in *vfr* may therefore be important for the adaptation of *P. aeruginosa* to the CF airways. We consequently hypothesized that this regulatory mutation in *vfr* was the first step of becoming a successful colonizer.

Here we investigated the point mutation in *vfr* and characterized regulatory consequences and pleiotropic effects caused by the mutation compared to wild type and a knockout mutant.

It was observed that the point mutation modified several phenotypes including transcriptional profile and catabolic performance towards a phenotype associated with chronic infection, indicating that a single mutation may cause completely alterations of multiple phenotypes reflecting the transition from an environmental strain to a host specific pathogen for CF patients.

Understanding the evolutionary paths by which *P. aeruginosa* adapts to the CF airways may provide useful information in the design of new treatment strategies for eradicating *P. aeruginosa*.

Results

Computational investigation of the vfr mutation

We have recently determined the sequence of genetic and phenotypic changes in the *P. aeruginosa* DK1 lineage that resulted in its successful establishment in the CF airways of a large number of patients from 1973 to 2012 [10]. The first mutation within a gene with regulatory function was an A-to-C transition at position 487 in the *vfr* gene. This mutation was found in DK1 isolates from 1973 and was immediately fixed in the lineage (i.e. found in all descending isolates including contemporary isolates).

The A487C mutation in *vfr* results in substitution of threonine-163 to a proline residue. SIFT analysis [26] predicted that this amino acid substitution had an effect of the function of the protein.

Based on the crystal structure of Vfr proposed by Cordes and colleagues [27] amino acids 163 and 164 have solvent-exposed side chains, and could therefore potentially participate in direct side-chain interactions required for transcriptional activation or DNA binding. Comparison of Vfr with its *E. coli* homolog CRP showed that the mutation occurred in the AR1 region (Figure 1). It has been reported that this region (AMTHPDG) is involved in CRP-RNA polymerase interaction in *E. coli* but not DNA binding and DNA bending [11, 14], indicating that the substitution of threonine-163 to proline could interfere with Vfr-RNA polymerase interaction and modify the function of Vfr. To verify this hypothesis, mutants of *vfr* was constructed and the phenotypic impact of the *vfr*^{T163P} mutant was examined and compared to a *vfr* knockout mutant.

Transcriptional profiles of vfr^{T163P} *and* Δ*vfr mutants*

We next sought to study the impact of the *vfr*^{T163P} mutation on gene expression. It has previously been reported that Vfr levels peak during late-exponential/early stationary phase [28, 29]. For this reason, transcriptome analysis was performed on PAO1, *vfr*^{T163P} and Δ*vfr* mutants grown to both exponential phase and late-exponential/early stationary phase in order to investigate the effect of Vfr on gene expression. The analysis was done using Affymetrix GeneChip (see Materials and methods). It was observed that under the experimental conditions, Vfr affected most genes during exponential phase. Of the 5883 genes and intergenic regions present on the microarray, 898 genes (15.3%) showed significantly altered expression in the *vfr*^{T163P} mutant compared to PAO1, and 1415 genes (24.1%) in the Δ*vfr* mutant compared to PAO1 ($P < 0.05$) at exponential phase (Table S1). In stationary phase, only 66 genes showed significantly altered expression in *vfr*^{T163P} mutant compared to PAO1, and 52 genes were differently expressed in Δ*vfr* mutant compared to PAO1 (Table S2), indicating that in stationary phase both mutants resembled wild type PAO1 compared with exponential phase data. Here, we therefore mainly focused on gene expression in exponential phase.

Venn diagrams were constructed to visualize the overlap and difference of gene expression in exponential phase between the mutants (Figure 2). From this it was observed that the two mutants overlapped with 589 differently expressed genes compared to PAO1. The Δ*vfr* mutant expressed 826 genes significantly different from the *vfr*^{T163P} mutant, while 309 genes were found to be differential expressed specific for the *vfr*^{T163P} mutant. From this it was clear that the two mutants were significantly different from PAO1 and that the *vfr*^{T163P} mutation indeed had functional effects. Importantly, the expression data showed that the two mutants were distinct which suggested that the *vfr*^{T163P} mutation modified Vfr and its regulon differently than a loss-of-function, knockout strain.

Mutations in vfr *result in CF associated phenotypes*

To experimentally address functional consequences of the observed transcriptome changes, we focused on differentially expressed genes with minimal twofold expression changes (Table 1).

Vfr has been described to be required for twitching motility [30]. Likewise, we observed that *pilJ* involved in twitching motility was downregulated in both mutants (Table 1). The Δ*vfr* mutant furthermore showed downregulation of the type 4 fimbrial precurser *pilA*. Measuring twitching zones in 1% agar plates confirmed this (Figure 3A). Whereas the *vfr* mutant significantly twitched less than PAO1, the Δ*vfr* mutant significantly twitched less than *vfr*^{T163P} (Student's t-test, $P < 0.05$), probably as a result of downregulation of *pilA*. In addition to its effect on twitching, Vfr has previously been described to negatively regulate the expression of flagellar genes by repressing the flagellar transcriptional regulator FleQ [20]. Dasgupta and colleagues found FleQ production to be downregulated following Vfr overproduction, but was unchanged in a Vfr knockout mutant [20]. Likewise, we did not observed difference in *fleQ* transcription in the *vfr*^{T163P} and the Δ*vfr* mutant. However, we did see a small but significant increase in swimming motility compared to PAO1 (Figure 3B), suggesting that even a small change in FleQ transcription may have an influence on swimming motility or that swimming motility may be regulated via a mechanism independent of FleQ. Both twitching and swimming motility was restored to wild-type levels in the complementary strain *vfr*-com.

Another group of genes was found to have differential expression in the *vfr* mutants compared to PAO1. The expression of genes involved in iron acquisition was increased in the *vfr*^{T163P} mutant including *pch* genes (required for production of the pyochelin sideophore) and *fptA* encoding the ferric-pyochelin receptor (Tabel 1). Furthermore, *ampP* (also known as *fptX*, involved in ferripyochelin transport), which has previously been described to be upregulated in iron-limited conditions, was found to be upregulated in the *vfr*^{T163P} mutant. This suggested that the *vfr*^{T163P} mutant had increased ability to utilize pyochelin and grow in iron-limited media. Indeed, when the *vfr*^{T163P} mutant and PAO1 were grown on iron-depleted VB plates supplemented with Fe^{3+} or Fe^{2+}, the *vfr*^{T163P} mutant grew significantly better than PAO1 (Student's t-test, *P*

161 < 0.05) (Figure 4A). Interestingly, the same effect was observed for the Δ*vfr* mutant although only slight
162 upregulation of genes involved in pyochelin production and reception was observed in this mutant (Table
163 1).
164 To verify that the growth under iron-limited condition was due to upregulation of pyochelin, we measured
165 the pyoverdin production. No difference was observed in pyoverdin production between the *vfr* mutants
166 and PAO1 (Figure 4B), suggesting that the ability to grow under iron limited condition was due to changed
167 pyochelin expression.
168 Vfr has previously been shown to directly bind and activate transcription of *lasR* [18]. LasR is the major
169 regulator of quorum sensing control in *P. aeruginosa* that positively regulates the expression of a large
170 number of genes including those for elastase (*lasB*) and the type II secretion apparatus. Therefore it was
171 not surprising that microarray data showed downreguation of *lasR* in both mutants (see below) and genes
172 belonging to type II secretion (Table S1). In agreement with this, protease secretion was reduced in both
173 mutants compared with PAO1 (Figure 5A). Another virulence factor described to be QS dependent was the
174 production of pyocyanin. The *vfr* mutants showed increased pyocyanin production compared to PAO1 after
175 16 hours of incubation (Figure 5B), indicating negative regulation of pyocyanin production probably
176 through *lasR*. This was supported by previous observations that a *lasR* mutant was found to have elevated
177 pyocyanin production compared to wild type after 14 hours [31].
178
179 *Mutations in* vfr *change catabolic performance*
180 The nutritional composition of the environment has an impact on evolution and adaptation of *P.*
181 *aeruginosa*. We therefore sought to examine the ability of the *vfr* mutants to utilize various carbon and
182 nitrogen sources using the Biolog Phenotype MicroArray (Materials and Methods). Significant differential
183 growth was observed on 38 substrates of both carbon and nitrogen sources ($P > 0.05$). The Biolog data
184 confirmed the observed difference between the *vfr*^{T163P} and the *vfr*::Gm mutant and further showed the
185 growth difference compared to PAO1 (Figure 6).
186 The *vfr* mutants were found to grow significantly better on several sugars including xylose, fucose, ribose,
187 and arabinose. Interestingly, both mutants showed significantly increased respiration of several amino
188 acids including proline, alanine and valine compared to PAO1. The differences in catabolic performance
189 between PAO1 and the *vfr* mutants on proline were supported by gene expression analysis showing
190 upregulation of the *proC* gene involved in proline metabolism (Table S1). In contrast, both mutants showed
191 significantly reduced respiration on nitrogen sources compared to PAO1. These data showed that Vfr
192 directly or indirectly regulated genes involved in metabolism and that mutations in Vfr changed the
193 catabolic performance of the strain.

194 *Differences between* vfr^{T163} *and the knockout mutants*

195 From the presented data it was clear that the two *vfr* mutants were significantly different from PAO1 in
196 expression of motility, iron acquisition, QS dependent phenotypes, and respiration of certain carbon and
197 nitrogen substrates. In contrast to this, we also observed differences between the *vfr*^{T163P} and the knockout
198 mutants. The *vfr*^{T163P} mutant utilized several nitrogen sources better than the *vfr*::Gm mutant including
199 alanine and serine, implying that under certain growth conditions *vfr*^{T163P} would grow better than the
200 *vfr*::Gm mutant.

201 Furthermore, microarray analysis revealed differences in gene expression between the two mutants not
202 only in exponential phase but also in stationary phase (Table 1), indicating that there were regulatory
203 difference between the *vfr*^{T163P} and the Δ*vfr* mutant. These differences were among others observed in the
204 expression of the *pqsA-E* operon, involved in the conversion of anthranilate to HHQ. The *pqsA-E* operon
205 was upregulated in Δ*vfr* compared to the *vfr*^{T163P} mutant (Table 1 and Table S1), suggesting that Vfr
206 negatively regulated the PQS system, however, this was not the case for *vfr*^{T163P} showing the same
207 transcriptional levels of the *pqs* operon as PAO1. This indicated that the modified Vfr from *vfr*^{T163P} is still
208 able to directly or indirectly regulate the *pqs* operon as in PAO1. In addition, both mutants showed the
209 metallic, iridescent sheen when grown on agar (Figure 7). This metallic sheen has been associated with the
210 accumulation of HHQ seen in *lasR* mutants [3]. However the *vfr*^{T163P} mutant exhibited a higher degree of
211 metallic sheen than the Δ*vfr* mutant (Figure 7). While both mutants downregulated *lasR* shown to
212 positively regulate *pqsH*, converting HHQ to PQS, Δ*vfr* showed 1.2 fold upregulation of *lasR* compared to
213 the *vfr*^{T163P} mutant. The difference could furthermore be seen in the expression of *pqsH*, which was
214 expressed 1.2 fold less in the *vfr*^{T163P} mutant compared to Δ*vfr*. This suggests that the observed difference
215 in metallic sheen may be caused by distinct rates of converting HHQ to PQS.

216

217 **Discussion**

218 Mutations in transcriptional regulators have important implications for *P. aeruginosa* when adapting to the
219 highly complex environment of CF patients. One mutation can have major phenotypic effects, as
220 transcriptional regulators often are central proteins in a complex regulatory network. Here we have
221 characterized the regulatory consequences and pleiotropic effects caused by a consistent *vfr* mutation
222 acquired during adaptation of the *P. aeruginosa* DK1 clone. The mutation (Thr163Pro) was within the RNA
223 polymerase interacting binding region AR1 (Figure 1). Single amino acid substitutions within AR1 of *E. coli*
224 CRP protein (Vfr homolog) have been demonstrated to reduce transcription activation but not affect DNA
225 binding and DNA bending [14]. In fact, Thr158 in CRP (Thr163 in Vfr) has been shown to be essential for
226 transcription activation and the most important side-chain for function of AR1 [32]. Furthermore, it was

proposed that Thr-158 made direct contact with RNA polymerase [32] and was therefore a very important site for interaction with RNA polymerase. We therefore argue that the observed mutation in *vfr* of clinical DK1 isolates modified the interaction with RNA polymerase resulting in re-directed regulatory capacity rather than loss-of-function. This hypothesis was supported by the presented data that showed differences in gene expression profiles, catabolic performance and colony morphology (i.e. metallic sheen) between *vfr*^{T163P} and knockout mutants. The single *vfr* mutation seemed to change the phenotype to a more chronic stage of phenotype represented by reduced protease secretion and twitching motility through downregulation of *lasR*, type II secretion genes, and pili genes. It has previously been reported that a *P. aeruginosa* PAK deficient in Vfr, did not have an effect on *pilA* transcription. However this seemed to be the case in a PAO1 background, indicating that there exist differences between the regulons of Vfr depending on the strain background

The *vfr*^{T163P} mutant was found to grow better on several amino acids including those found in high concentrations in sputum samples. The CF sputum has previously been reported to have increased availability of amino acid compared to healthy individuals [2], and clinical isolates have been shown to use these amino acids as preferred carbon sources in sputum [33], indicating that *vfr* mutants may have an selective advantage in the CF airways. The *vfr*^{T163P} mutant utilized several nitrogen sources better than the Δ*vfr* mutant including alanine and serine, implying that the point mutation would provide a selective advantage in the CF habitat. In addition, we observed that the *vfr* mutants grew significantly better under iron limited condition compared to wild type PAO1. The transcriptional regulator PchR regulates expression of the *pchABCDE* and *fptA* genes, however *pchR* was not found to be upregulated, indicating that the modified Vfr produced by the *vfr*^{T163P} mutant may function as an alternate regulator of the *pch* operon or that the increase in expression level of the *pch* operon was due to change in mRNA stability rather than transcription rate.

Even though there was a difference in upregulation of *pch* genes between *vfr*^{T163P} and Δ*vfr*, this difference was not detected in the iron depletion assay. It should be kept in mind that the conditions in which gene expression and iron depletion were measured were completely different. Therefore it is possible that under the iron depletion condition both *vfr* mutants show increased expression of pyochelin compared to PAO1.

Previously, it has been documented that *P. aeruginosa* grows under iron-restricted conditions in the CF airways [34], supporting the hypothesis that mutations in *vfr* would confer a selective advantage when adapting to the CF environment. This was further supported by previous observations by Fox and colleagues showing that under certain growth conditions there was a positive selection of *vfr* mutants and demonstrated that *vfr* mutants had a selective advantage over wild type after growth in static cultures [35]. These results are quite intriguing as *vfr* mutants of *P. aeruginosa* naturally and frequently arise in the cystic

fibrosis environment during long-term colonization [36]. We therefore propose that this single point mutation in *vfr* was the first step of becoming a successful colonizer of the CF airways, giving *P. aeruginosa* multiple selective advantages.

The expression of other known Vfr targets such as *regA*, *fleQ*, *ptxR* and *toxA* were not altered in the Vfr mutants indicating that under the investigated conditions numerous other regulators may control transcription of these genes.

Experimental evolution studies conducted in laboratory conditions have frequently documented that transcriptional regulators are targets for adaptive mutations [8, 9]. Recently, this was confirmed for a number of regulatory mutations observed to arise in the DK2 clone These mutations helped the DK2 clone to adapt by continuous optimization of existing regulatory networks [7]. Likewise, we identified an early appearing mutation in the global transcriptional regulator Vfr in the DK1 lineage. This mutation was observed in all DK1 isolates from 1973 and onward with the exception of one historical isolate, that surprisingly enough had a mutation in *lasR* instead, indicating that the two mutations might compensate for each other. This was supported by the observation that *vfr* mutants shared many of the phenotypes of a *lasR* mutant, indicating that the same phenotypic outcome is possible by different evolutionary paths.

Based on the presented data, we therefore propose that this single point mutation in *vfr* was the first step of becoming a successful colonizer of the CF airways, giving *P. aeruginosa* multiple selective advantages.

In conclusion, vfr^{T163P} was shown to be a mutant of PAO1 distinct from a knockout mutant and was found to have great pleiotropic effects on gene expression and catabolic performance.

Method

Bacterial strains and plasmids

The bacterial strains used in this study are described in Table 2. *E. coli* and *P. aeruginosa* strains were routinely grown in Luria-Bertani (LB) broth or on LB agar at 37 °C. When needed appropriate antibiotic was added.

Construction of vfr mutants

The *vfr*^{T163P} mutant and the *vfr* deletion mutant (Δ*vfr*) was constructed by allelic replacement. The *vfr* gene containing the mutation *vfr*(A487T) was amplified from a DK1 isolate using the primers vfr-fw-SacI (GAGAGAGCTCCAGACTCCCGCCGGCAAAGG) and vfr-rev-XbaI (ACATTCTAGAGGACACGCCGTTGATGGGCA) or vfr-rev-PstI (ACATCTGCAGGGACACGCCGTTGATGGGCA). The fragment was digested with SacI and XbaI/PstI and ligated onto the plasmid pNJ1 and transformed into *E. coli* CC118λpir by standard protocols [37].

The deletion mutant Vfr-null was constructed by cutting out two base pair at position 108 in the *vfr* gene on pTM1 creating a frame shift. The insertion mutant TM4 was constructed by inserting a gentamycin cassette at position 233.

The mutants were constructed by triparental matings in which the recipient received pTM1 or pTN2. *E. coli* HB101 (pRK600) was used as helper strain and *E. coli* CC118λpir as donor.

A complement strain of *vfr*^{T163P} was constructed by introducing plasmid pME6031 harboring the wt *vfr* gene (pTN1) by electroporation.

Affymetrix GeneChip

Strains were grown aerobically at 37°C in LB medium starting at OD_{600} = 0.01 and harvested at OD_{600} = 0.5 and OD_{600} = 0.8. Immediately after harvesting, cells were mixed with STOP solution (95% ethanol, 5% phenol), pelleted by centrifugation and stored at -80°C. Total RNA was extracted by RNeasy Mini Kit (Qiagen, Germany) and the RNase-Free DNase Set was used for a more complete DNA removal. The total yield of RNA and A_{260}/A_{280} was measured using NanoDrop. Preparation of cDNA, labelling, hybridization and processing of the *P. aeruginosa* GeneChip were performed at the RH Microarray Center, Department for Genomic Medicine, Rigshospitalet, Copenhagen University. The raw CEL files were obtained using the Affymetrix GeneChip Operating System 1.4 and analyzed using BioConductor for the statistical software R. Strains were tested in triplicates.

Biolog Phenotype MicroArray™

The catabolic performance of the *vfr* mutants and reference strain PAO1 was tested on carbon and nitrogen sources by the Biolog Phenotype MicroArrays (Biolog, Hayward, CA). The Biolog Phenotype Microarrays were performed in duplicates according to manufacturer instructions [38, 39]. Briefly, strains were streaked on LB agar plates and incubated at 37°C for 24 h. Cells were swabbed from the plates and suspended in IF-0 GN Base (inoculation fluid) at a density corresponding to 42% transmittance in the Biolog turbidimeter. The cell suspensions were diluted 1:6 in IF-0 minimal medium containing Biolog redox dye mixture D (tetrazolium), and 100 µL aliquots were added to carbon source plates (PM1). Nitrogen plates (PM3) were prepared with ferric citrate and glucose supplemented to the inoculation mixture. In total 190 substrates were available on the two plates. Of these 126 were found to be informative by showing consistent growth between replicates and not autoinducing the color change (49 carbon substrates and 77 nitrogen sources). OmniLog *OL_FM/Kin* 1.20.02 software (Biolog) was used to export of the OmniLog data and the average area beneath each kinetic curve was used for analysis. Further data and statistical analysis was performed using R version 2.10.1 (R Development Core Team 2009).

Motility assay

The swimming and twitching ability of each mutant was measured on ABT plates (AB medium containing 2.5 mg/L thiamine) supplemented with 0.5% casamino acids and 0.5% glucose. For swimming motility, colonies grown overnight (ON) on LB agar were stabbed in the middle of the 0.3% (wt/vol) ABT agar using a sterile toothpick and incubated 24 hours at 30°C. For twitching motility, a colony of each isolate was stabbed though 1.5% (wt/vol) ABT agar and placed on the bottom. Twitching motility was determined by measuring the diameters of the twitching zone after 48 hours of incubation at 37°C.

The determination of swimming and twitching motility was performed in triplicates and the motility zones were calculated relative to the reference strain PAO1.

Determination of pyocyanin production

The pyocyanin concentration produced by PAO1, TM1 and vfr-null was determined as previously described [40]. Briefly, culture supernatant was mixed with 0.3 vol of chloroform. The organic phase was removed and acidified by addition of 0.1 vol 0.2 M HCl. The pink top layer was removed and its absorbance was measured at 520 nm. The OD_{520} was multiplied with 17.072 to get the pyocyanin concentration in µg/mL.

Skim milk protease assay

The production of protease was determined by applying one colony of the isolate grown on LB ON to skim milk plates (LB agar with 10% skim milk). The plates were incubated for 24 h at 37°C and the clearing zones were measured.

Iron depletion assay

Bacteria grown ON in LB broth were repeatedly washed with 0.9% (wt/vol) NaCl, and 100 µL of OD_{600} were plated on iron-depleted Vogel-Bonner minimal medium agar (200 mg/L $MgSO_4 \bullet 7H_2O$, 2 g/L citric acid monohydrate, 10 g/L K_2HPO_4, 3.5 g/L $NaNH_4PO_4$, 2% glucose) supplemented with 2 mM iron chelator 2,2-dipyridyl. Sterile filter disks were placed on top of the agar and 8 µL Fe^{2+} (200 mM ferrous sulfate) or Fe^{3+} (100 mM ferric citrate) and incubated at 37°C for 24h.

Pyoverdin measurements

Pyoverdin concentration was determined as previously described [41]. Briefly, strains were grown overnight in ABT minimal medium containing 0.01 $FeCl_3$ plus 0.5% casamino acids supplemented with 50 µM 2,2-dipyridyl to induce pyoverdine production. The supernatant was diluted 10X in 1 M of Tris-HCl (pH 8) and the pyoverdin concentration was determined as OD_{405} normalized with OD_{600}.

Acknowledgement

We would like to thank S. Damkiær for assistance with mutant construction. We would also like to thank the Infection Microbiology Group for helpful discussions.

This work is supported by the Danish Counsil for Independent Research and the Lundbeck Foundation.

Author contributions: T.M.M., S.M and L.J. conceived the study and designed research; T.M.M. and T.N. performed research; T.M.M., T.N., and L.J. analysed the data and T.M.M. and L.J. wrote the manuscript.

References

1. Jensen, E.T., et al., *Epidemiology of Pseudomonas aeruginosa in cystic fibrosis and the possible role of contamination by dental equipment.* The Journal of hospital infection, 1997. **36**(2): p. 117-22.
2. Barth, A.L. and T.L. Pitt, *The high amino-acid content of sputum from cystic fibrosis patients promotes growth of auxotrophic Pseudomonas aeruginosa.* Journal of medical microbiology, 1996. **45**(2): p. 110-9.
3. D'Argenio, D.A., et al., *Growth phenotypes of Pseudomonas aeruginosa lasR mutants adapted to the airways of cystic fibrosis patients.* Mol Microbiol, 2007. **64**(2): p. 512-33.
4. Deretic, V., M.J. Schurr, and H. Yu, *Pseudomonas aeruginosa, mucoidy and the chronic infection phenotype in cystic fibrosis.* Trends in microbiology, 1995. **3**(9): p. 351-6.

5. Mahenthiralingam, E., M.E. Campbell, and D.P. Speert, *Nonmotility and phagocytic resistance of Pseudomonas aeruginosa isolates from chronically colonized patients with cystic fibrosis.* Infect Immun, 1994. **62**(2): p. 596-605.
6. Wang, L., F.F. Wang, and W. Qian, *Evolutionary rewiring and reprogramming of bacterial transcription regulation.* Journal of genetics and genomics = Yi chuan xue bao, 2011. **38**(7): p. 279-88.
7. Damkiaer, S., et al., *Evolutionary remodeling of global regulatory networks during long-term bacterial adaptation to human hosts.* Proc Natl Acad Sci U S A, 2013. **110**(19): p. 7766-71.
8. Carroll, S.B., *Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution.* Cell, 2008. **134**(1): p. 25-36.
9. Wang, L., et al., *Divergence involving global regulatory gene mutations in an Escherichia coli population evolving under phosphate limitation.* Genome biology and evolution, 2010. **2**: p. 478-87.
10. Markussen, T., et al., *Niche partitioning results in diversification and long-term co-existence of Pseudomonas aeruginosa sub-lineages during chronic CF airway infections.* Submitted, 2014.
11. West, S.E., A.K. Sample, and L.J. Runyen-Janecky, *The vfr gene product, required for Pseudomonas aeruginosa exotoxin A and protease production, belongs to the cyclic AMP receptor protein family.* J Bacteriol, 1994. **176**(24): p. 7532-42.
12. Busby, S. and R.H. Ebright, *Transcription activation by catabolite activator protein (CAP).* J Mol Biol, 1999. **293**(2): p. 199-213.
13. Jin, R., K.A. Sharif, and J.S. Krakow, *Evidence for contact between the cyclic AMP receptor protein and the delta 70 subunit of Escherichia coli RNA polymerase.* J Biol Chem, 1995. **270**(33): p. 19213-6.
14. Zhou, Y., X. Zhang, and R.H. Ebright, *Identification of the activating region of catabolite gene activator protein (CAP): isolation and characterization of mutants of CAP specifically defective in transcription activation.* Proc Natl Acad Sci U S A, 1993. **90**(13): p. 6081-5.
15. West, D., et al., *Interactions between the Escherichia coli cyclic AMP receptor protein and RNA polymerase at class II promoters.* Mol Microbiol, 1993. **10**(4): p. 789-97.
16. Suh, S.J., et al., *Effect of vfr mutation on global gene expression and catabolite repression control of Pseudomonas aeruginosa.* Microbiology, 2002. **148**(Pt 5): p. 1561-9.
17. VanderLeest, H.T., et al., *Seasonal encoding by the circadian pacemaker of the SCN.* Curr Biol, 2007. **17**(5): p. 468-73.
18. Albus, A.M., et al., *Vfr controls quorum sensing in Pseudomonas aeruginosa.* J Bacteriol, 1997. **179**(12): p. 3928-35.
19. Fuchs, E.L., et al., *The Pseudomonas aeruginosa Vfr regulator controls global virulence factor expression through cyclic AMP-dependent and -independent mechanisms.* J Bacteriol, 2010. **192**(14): p. 3553-64.
20. Dasgupta, N., et al., *fleQ, the gene encoding the major flagellar regulator of Pseudomonas aeruginosa, is sigma70 dependent and is downregulated by Vfr, a homolog of Escherichia coli cyclic AMP receptor protein.* J Bacteriol, 2002. **184**(19): p. 5240-50.
21. Ferrell, E., et al., *Regulation of Pseudomonas aeruginosa ptxR by Vfr.* Microbiology, 2008. **154**(Pt 2): p. 431-9.
22. Kanack, K.J., et al., *Characterization of DNA-binding specificity and analysis of binding sites of the Pseudomonas aeruginosa global regulator, Vfr, a homologue of the Escherichia coli cAMP receptor protein.* Microbiology, 2006. **152**(Pt 12): p. 3485-96.
23. Fuchs, E.L., et al., *In vitro and in vivo characterization of the Pseudomonas aeruginosa cyclic AMP (cAMP) phosphodiesterase CpdA, required for cAMP homeostasis and virulence factor regulation.* J Bacteriol, 2010. **192**(11): p. 2779-90.
24. Whitchurch, C.B., et al., *Pseudomonas aeruginosa fimL regulates multiple virulence functions by intersecting with Vfr-modulated pathways.* Mol Microbiol, 2005. **55**(5): p. 1357-78.

25. Croda-Garcia, G., et al., *Transcriptional regulation of Pseudomonas aeruginosa rhlR: role of the CRP orthologue Vfr (virulence factor regulator) and quorum-sensing regulators LasR and RhlR.* Microbiology, 2011. **157**(Pt 9): p. 2545-55.
26. Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function.* Nucleic Acids Res, 2003. **31**(13): p. 3812-4.
27. Cordes, T.J., et al., *Crystal structure of the Pseudomonas aeruginosa virulence factor regulator.* J Bacteriol, 2011. **193**(16): p. 4069-74.
28. Burrowes, E., et al., *Influence of the regulatory protein RsmA on cellular functions in Pseudomonas aeruginosa PAO1, as revealed by transcriptome analysis.* Microbiology, 2006. **152**(Pt 2): p. 405-18.
29. Heurlier, K., et al., *Negative control of quorum sensing by RpoN (sigma54) in Pseudomonas aeruginosa PAO1.* J Bacteriol, 2003. **185**(7): p. 2227-35.
30. Beatson, S.A., et al., *Differential regulation of twitching motility and elastase production by Vfr in Pseudomonas aeruginosa.* J Bacteriol, 2002. **184**(13): p. 3605-13.
31. Diggle, S.P., et al., *Advancing the quorum in Pseudomonas aeruginosa: MvaT and the regulation of N-acylhomoserine lactone production and virulence gene expression.* J Bacteriol, 2002. **184**(10): p. 2576-86.
32. Niu, W., et al., *Characterization of the activating region of Escherichia coli catabolite gene activator protein (CAP). I. Saturation and alanine-scanning mutagenesis.* J Mol Biol, 1994. **243**(4): p. 595-602.
33. Palmer, K.L., L.M. Aye, and M. Whiteley, *Nutritional cues control Pseudomonas aeruginosa multicellular behavior in cystic fibrosis sputum.* J Bacteriol, 2007. **189**(22): p. 8079-87.
34. Haas, B., et al., *Siderophore presence in sputa of cystic fibrosis patients.* Infect Immun, 1991. **59**(11): p. 3997-4000.
35. Fox, A., et al., *Emergence of secretion-defective sublines of Pseudomonas aeruginosa PAO1 resulting from spontaneous mutations in the vfr global regulatory gene.* Appl Environ Microbiol, 2008. **74**(6): p. 1902-8.
36. Smith, E.E., et al., *Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients.* Proc Natl Acad Sci U S A, 2006. **103**(22): p. 8487-92.
37. Finch, L., H. Barbeau, and B. Arsenault, *Influence of body weight support on normal human gait: development of a gait retraining strategy.* Physical therapy, 1991. **71**(11): p. 842-55; discussion 855-6.
38. Bochner, B.R., *New technologies to assess genotype-phenotype relationships.* Nat Rev Genet, 2003. **4**(4): p. 309-14.
39. Bochner, B.R., *Global phenotypic characterization of bacteria.* FEMS Microbiol Rev, 2009. **33**(1): p. 191-205.
40. Essar, D.W., et al., *Identification and characterization of genes for a second anthranilate synthase in Pseudomonas aeruginosa: interchangeability of the two anthranilate synthases and evolutionary implications.* J Bacteriol, 1990. **172**(2): p. 884-900.
41. Imperi, F., F. Tiburzi, and P. Visca, *Molecular basis of pyoverdine siderophore recycling in Pseudomonas aeruginosa.* Proc Natl Acad Sci U S A, 2009. **106**(48): p. 20440-5.
42. Herrero, M., V. de Lorenzo, and K.N. Timmis, *Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria.* J Bacteriol, 1990. **172**(11): p. 6557-67.
43. Kessler, B., V. de Lorenzo, and K.N. Timmis, *A general system to integrate lacZ fusions into the chromosomes of gram-negative eubacteria: regulation of the Pm promoter of the TOL plasmid studied with all controlling elements in monocopy.* Mol Gen Genet, 1992. **233**(1-2): p. 293-301.
44. Holloway, B.W., V. Krishnapillai, and A.F. Morgan, *Chromosomal genetics of Pseudomonas.* Microbiol Rev, 1979. **43**(1): p. 73-102.
45. Marvig, R.L., et al., *Within-host evolution of Pseudomonas aeruginosa reveals adaptation towards iron acquisition from hemoglobin.* Submitted, 2013.

46. Yang, L., et al., *Polysaccharides serve as scaffold of biofilms formed by mucoid Pseudomonas aeruginosa.* FEMS immunology and medical microbiology, 2012. **65**(2): p. 366-76.
47. Heeb, S., et al., *Small, stable shuttle vectors based on the minimal pVS1 replicon for use in gram-negative, plant-associated bacteria.* Molecular plant-microbe interactions : MPMI, 2000. **13**(2): p. 232-7.

Figures:

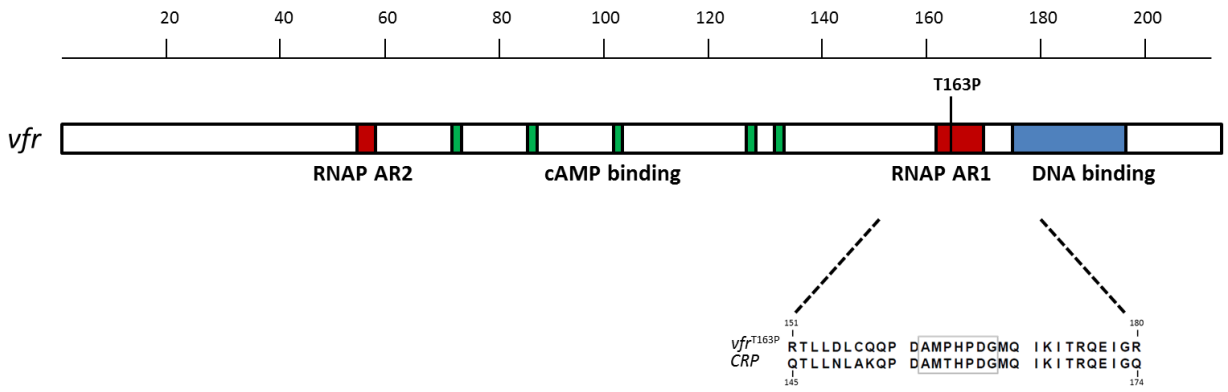


Figure 1. Schematic overview of sequence annotation of the Vfr protein. RNA polymerase interaction sites AR1 and AR2 are represented in red, cAMP binding sites are shown in green, and DNA binding sites are shown in blue. The location of the observed *vfr* mutation is marked in the AR1 region and aligned with the *E. coli* CRP protein.

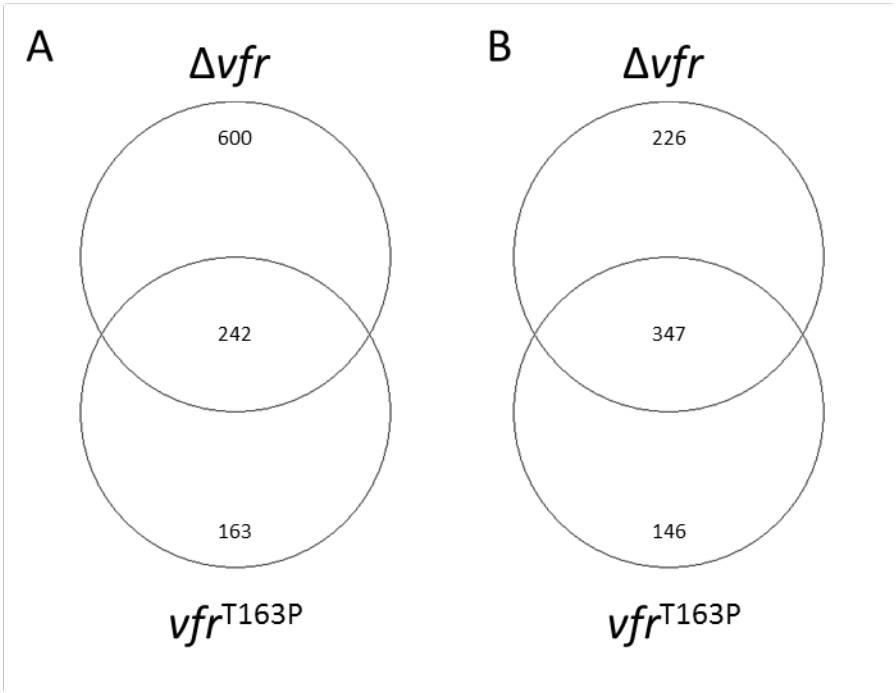


Figure 2. Venn diagram showing significantly ($P < 0.05$) upregulated genes of Δ*vfr* and the *vfr*^{T163P} mutants compared to PAO1 (A), and downregulated genes of Δ*vfr* and the *vfr*^{T163P} mutants compared to PAO1 (B).

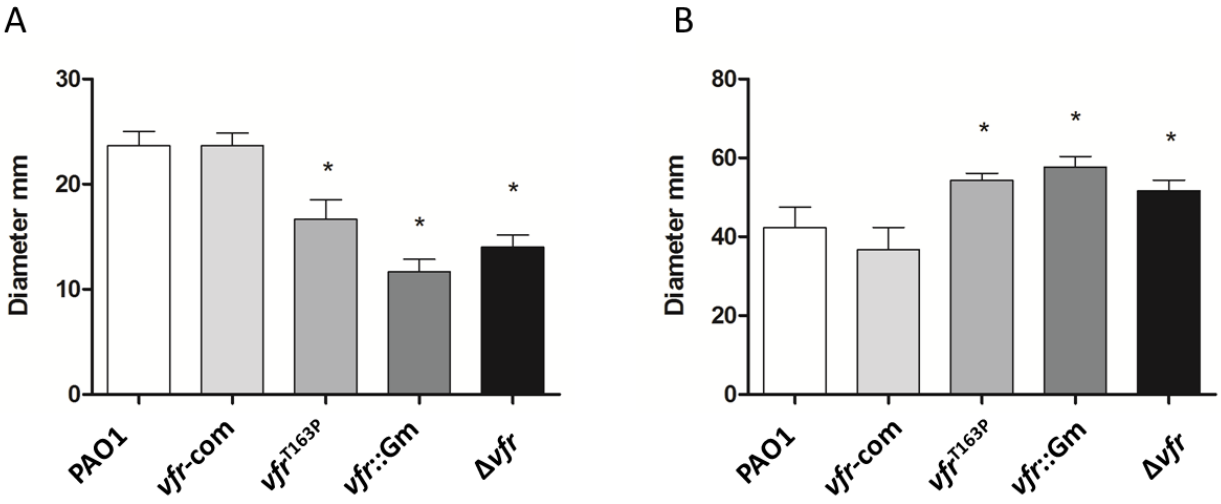


Figure 3. Motility diameters of PAO1 and *vfr* mutants. A) Twitching motility. Significant reduction was observed between PAO1 and all of the Vfr mutants. B) Swimming motility. All mutants showed significantly increased swimming compared with PAO1.

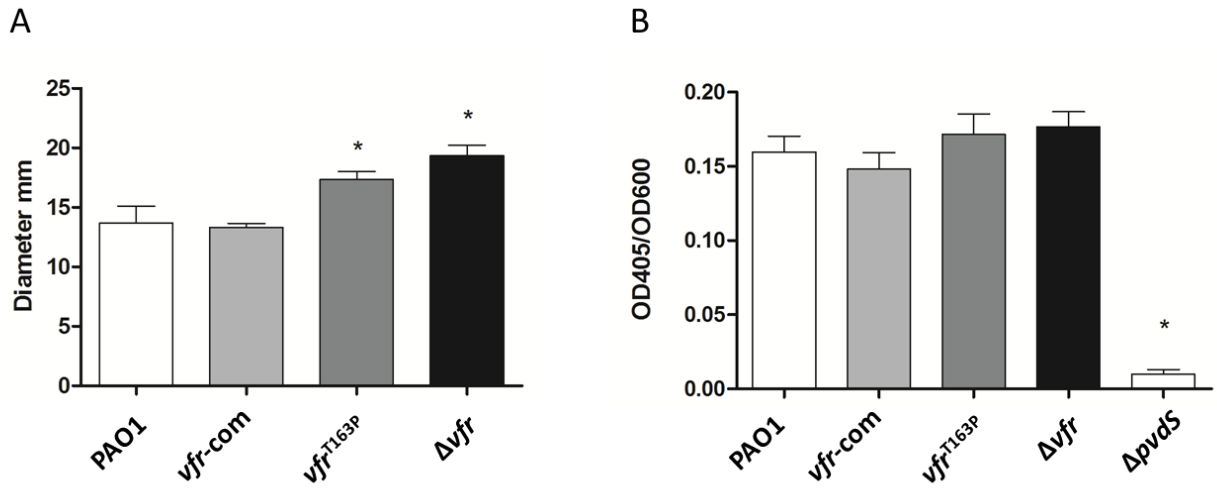
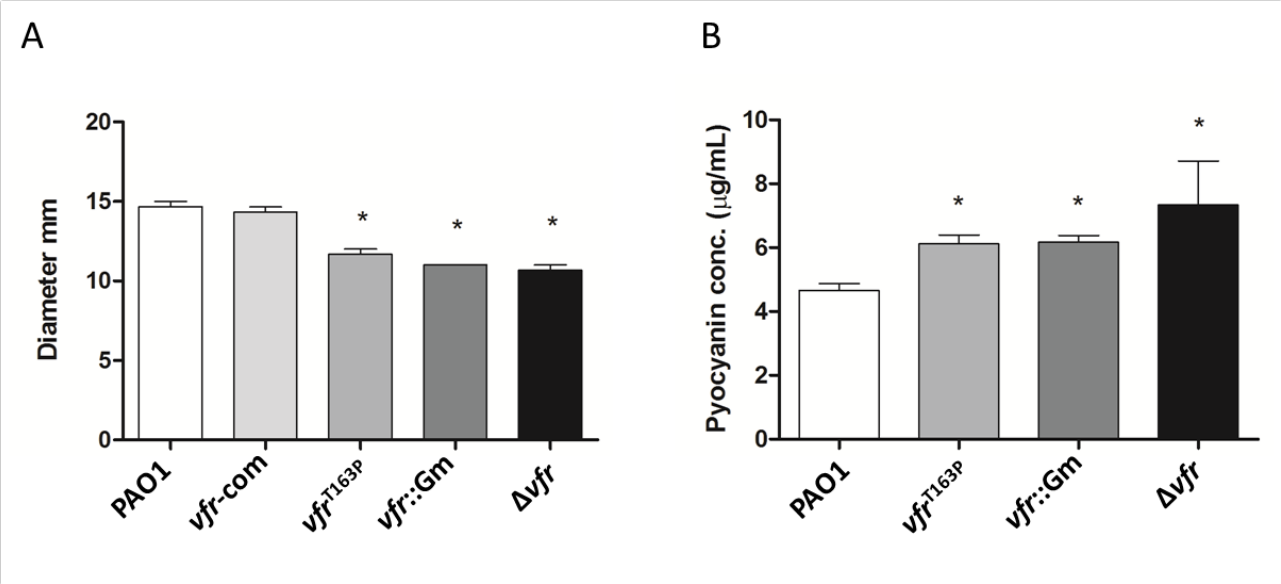


Figure 4. Iron acquisition of mutants. A) Growth zones of PAO1, vfr-com, vfr^{T163P}, Δ*vfr* under iron limited conditions. Bacteria were spread on iron depleted VB plates and repleated with Fe^{2+} (200 mM) on applied filter disks. The same result was observed when applying Fe^{3+} (data not shown). B) Pyoverdin concentration of PAO1, vfr-com, the vfr^{T163P} mutant, and Δ*vfr*. Knockout mutant Δ*pvdS* of PAO1 was used as negative control.



508

509 **Figure 5. Pyocyanin production in mutant strains.** The *vfr*^{T163P}, *vfr*::Gm, and Δ*vfr* mutants were significant
510 different from PAO1, while no significant difference was observed between the isolates (Student's t-test, *P*
511 < 0.05).

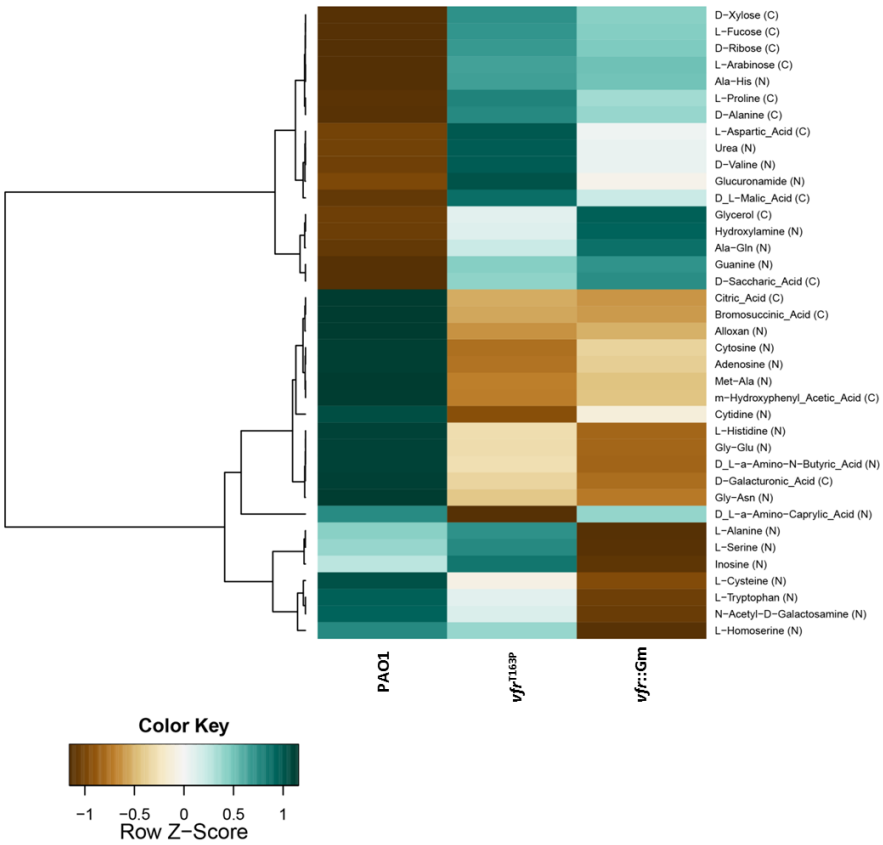


Figure 6. Analysis of catabolic performance of PAO1, the *vfr*^{T163P} mutant, and *vfr*::Gm represented as a heatmap. Significant differences in catabolic performance among PAO1, the *vfr*^{T163P} mutant and Δ*vfr* were observed for 38 substrates (student's t-test, $P < 0.05$).

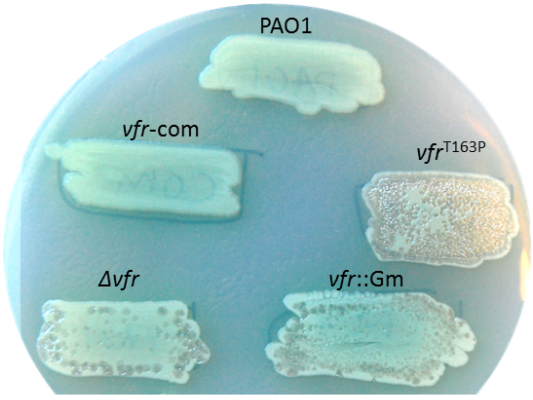


Figure 7. Vfr mutants show variant degree of metallic sheen. Growth on LB agar, photographed after incubation overnight at 37°C.

521

522 TABLES:

523 **Table 1. Gene expression of the *vfr*^{T163P} mutant and Δ*vfr* compared to *P. aeruginosa* PAO1.**

Gene identity	Gene name	Gene_product_name	Exponential phase			Stationary phase		
			PAO1/ *vfr*^{T163P} [a]	PAO1/Δ*vfr* [a]	*vfr*^{T163P} [b] /Δ*vfr*	PAO1/ *vfr*^{T163P}	PAO1/Δ*vfr*	*vfr*^{T163P} /Δ*vfr*
PA4218	*ampP*	AmpP	**2,26**	1,16	-	-	-	-
PA4220	*fptB*	hypothetical protein	**4,56**	1,38	**-3,30**	-	-	-
PA4221	*fptA*	Fe(III)-pyochelin outer membrane receptor precursor	**3,76**	1,23	**-3,05**	-	-	-
PA4224	*pchG*	pyochelin biosynthetic protein PchG	**2,03**	1,20	-1,70	-	-	-
PA4226	*pchE*	dihydroaeruginoic acid synthetase	**2,77**	1,20	**-2,31**	-	-	-
PA4228	*pchD*	pyochelin biosynthesis protein PchD	**2,20**	1,11	-1,98	-	-	-
PA4229	*pchC*	pyochelin biosynthetic protein PchC	**2,62**	1,25	**-2,11**	-	-	-
PA4230	*pchB*	salicylate biosynthesis protein PchB	**2,59**	-	**-2,14**	-	-	-
PA3397	*fpr*	ferredoxin--NADP+ reductase	**-2,10**	-1,88	**-**	-	-	-
PA3476	*rhlI*	autoinducer synthesis protein RhlI	**-2,35**	-1,56	1,51	-	-	-
PA3904		hypothetical protein	**-3,64**	-1,75	**2,07**	-	-	-
PA3905		hypothetical protein	**-2,01**	-	1,47	-	-	-
PA3907		hypothetical protein	**-2,20**	-	1,48	-	-	-
PA3908		hypothetical protein	**-2,04**	-	**-**	-	-	-
PA4778	*cueR*	CueR	**-2,02**	-1,49	1,35	-	-	-
PA0997	*pqsB*	PqsB	-	**2,04**	**2,92**	-	-	-
PA0411	*pilJ*	twitching motility protein PilJ	-1,95	**-2,19**	**-**	-	-	-
PA0652	*vfr*	transcriptional regulator Vfr	-1,73	**-2,69**	-1,55	-	-	-
PA3266	*capB*	cold acclimation protein B	-1,82	**-2,13**	**-**	-	-	-
PA4525	*pilA*	type 4 fimbrial precursor PilA	-	**-2,28**	**-**	-	-	-
PA5240	*trxA*	thioredoxin	-1,83	**-2,20**	**-**	-	-	-
PA0026	*plcB*	phospholipase C, PlcB	**-2,07**	**-2,58**	**-**	-	-	-
PA0563		conserved hypothetical protein	**-2,11**	**-2,21**	**-**	-	-	-
PA2761		hypothetical protein	**-2,24**	**-2,36**	**-**	-	-	-
PA3031		hypothetical protein	**-2,01**	**-2,23**	**-**	-	-	-
PA3713	*spdH*	spermidine dehydrogenase, SpdH	**-2,67**	**-2,82**	**-**	-	-	-
PA4616		probable c4-dicarboxylate-binding protein	**-2,63**	**-2,83**	**-**	-	-	-
PA4747	*secG*	secretion protein SecG	**-2,53**	**-2,27**	**-**	-	-	-
PA5049	*rpmE*	50S ribosomal protein L31	**-2,25**	**-2,42**	**-**	-	-	-
PA0996	*pqsA*	probable coenzyme A ligase	-	-	**2,04**	-	-	-
PA0998	*pqsC*	PqsC	-	-	**2,92**	-	-	-
PA0999	*pqsD*	3-oxoacyl-[acyl-carrier-protein] synthase III	-	-	**2,65**	-	-	-
PA0266	*gabT*	4-aminobutyrate aminotransferase	-	-	-	**2,60**	**2,04**	-1,28
PA0446		conserved hypothetical protein	-	-	-	**3,01**	-	-
PA3569	*mmsB*	3-hydroxyisobutyrate dehydrogenase	-	-	-	**4,03**	-	-
PA3570	*mmsA*	methylmalonate-semialdehyde dehydrogenase	-	-	-	**4,30**	-	-
PA1342		probable binding protein component of ABC transporter	-	-	-	**-2,44**	-	-
PA5507		hypothetical protein	-	-	-	**-5,55**	-	-
PA0887	*acsA*	acetyl-coenzyme A synthetase	-	-	-	-	**-9,61**	-

PA5083	conserved hypothetical protein	-	-	-	-	**4,88**	-
PA5082	probable binding protein component of ABC transporter	-	-	-	-	**5,65**	-

[a] Positive values represent upregulation of mutant gene expression compared to PAO1, while negative values represent downregulation of mutant gene expression compared to PAO1.

[b] Positive values represent upregulation in the Δ*vfr* mutant compared to the *vfr*^{T163P} mutant, while negative values represent downregulation in Δ*vfr* compared to the *vfr*^{T163P} mutant. The list does not include tRNA genes or intergenic regions. For full list see Table S1 and S2.

Table 2: Bacterial strains and plasmids used in this study.

Strains, plasmids or primers	Relevant genotype or phenotype	Source of reference
Strains		
E. coli		
CC118λpir	*E.coli* CC18 lysogenized with lambda pir phage	[42]
HB101	*recA,* thi, pro, leu, hsd RM^{+}; Sm^{R}	[43]
P. aeruginosa		
PAO1	Wild-type	[44]
vfr^{T163P}	PAO1 *vfr*(A487T)	This study
vfr::Gm	PAO1 Δ*vfr* , Gm^{r}	This study
Δ*vfr*	PAO1 Δ*vfr*	This study
vfr-com	*vfr*^{T163P} complemented *in trans* by electroporation with pTN1	This study
Δ*pvdS*	PAO1 Δ*pvdS*	[45]
Plasmids		
pNJ1	Tc^{R}, *sacB*$^{+}$ R6K *ori*, mobRp4, allelic replacement vector derived from pDS132	[46]
pME6031	Tc^{R}	[47]
pTM1	, pNJ1 with *SacI-XbaI* fragment containing PAO1 *vfr* (A487T)	This study
pTN1	, pME6031 with *SacI-XbaI* fragment containing PAO1 wt *vfr*	This study
pTM4	, pNJ1 with *SacI-XbaI* fragment containing PAO1 *vfr* deleted by Gm^{R} cassette	This study
pTN2	, pNJ1 with *SacI-PstI* fragment containing PAO1 *vfr* with deletion of two base pairs resulting in a frame shift.	This study
pRK600	Cm^{R}, *ori* ColE1 RK2-Mob^{+} RK2-Tra^{+} helper plasmid in triparental matings	[43]

Supplementary information

Table S1.

https://www.dropbox.com/sh/m59pwhg91edn80z/JgZnkeRDqO

Table S2. Gene expression changes in vfr-point and vfr-del compared to PAO1 in stationary phase.

Gene	Gene_name	Gene_product_name	PAO1/*vfr*T163P	PAO1/Δ*vfr*	*vfr*T163P/Δ*vfr*
PA5507		hypothetical protein	-5,55	-	-
PA1342		probable binding protein component of ABC transporter	-2,44	-	-
PA4211	phzB1	probable phenazine biosynthesis protein	-1,80	-	2,17
PA4141		hypothetical protein	-1,70	-	-
PA5555	atpG	ATP synthase gamma chain	-1,70	-	-
PA4130		probable sulfite or nitrite reductase	-1,60	-	-
PA4762	grpE	heat shock protein GrpE	-1,49	-	-
PA1902	phzD2	phenazine biosynthesis protein PhzD	-1,35	-	-
PA1999	dhcA	DhcA, dehydrocarnitine CoA transferase, subunit A	-1,28	-	-
PA3660	yjcE	probable sodium/hydrogen antiporter	-1,23	-	-
PA5295		hypothetical protein	-1,18	-	-
PA4547	pilR	two-component response regulator PilR	-1,17	-	-
PA4013	yohK	conserved hypothetical protein	-1,17	-1,24	-
PA3759		probable aminotransferase	-1,16	-	-
PA0076	tagF1	TagF1	-1,15	-	-
PA3284		hypothetical protein	-1,15	-	-
PA0230	pcaB	3-carboxy-cis,cis-muconate cycloisomerase	-1,13	-	-
PA1209		hypothetical protein	-1,13	-1,10	-
PA5486	yhgN	conserved hypothetical protein	-1,12	-	-
PA3850		hypothetical protein	-1,12	-	-
PA1491		probable transporter	-1,12	-	-
PA1693	pscR	translocation protein in type III secretion	-1,11	-	-
PA0360		hypothetical protein	-1,11	-	-
PA3670		hypothetical protein	-1,09	-	-
PA4953	motB	chemotaxis protein MotB	-1,08	-	-
PA1067		probable transcriptional regulator	-1,08	-1,03	-
PA0981		hypothetical protein	-1,07	-	-
PA4369		hypothetical protein	-1,07	-	-
PA0415	chpC	probableChemotaxis protein	-1,06	-	-
PA0568		hypothetical protein	1,01	-	-
PA4533		hypothetical protein	1,04	-1,10	-1,14
PA4378	inaA	InaA protein	1,04	-	-
PA1595		hypothetical protein	1,06	-	-
PA5369	pstS	phosphate ABC transporter, periplasmic phosphate-binding protein, PstS	1,06	-	-

PA1565		probable oxidoreductase	1,06	-	-
PA2404		hypothetical protein	1,07	-	-
PA0534		conserved hypothetical protein	1,07	-	-
PA1182		probable transcriptional regulator	1,08	-	-
PA5309		probable oxidoreductase	1,10	-	-
PA1433		conserved hypothetical protein	1,10	-	-
PA2525	opmB	probable outer membrane protein precursor	1,11	-	-
PA1568		conserved hypothetical protein	1,12	-	-
PA2584	pgsA	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	1,12	-	-1,06
PA1191		hypothetical protein	1,13	-	-
PA1107		conserved hypothetical protein	1,16	-	-1,16
PA1788		hypothetical protein	1,16	-	-
PA1919	nrdG	class III (anaerobic) ribonucleoside-triphosphate reductase activating protein, 'activase', NrdG	1,17	-	-
ig_3265210_3265847	NA	NA	1,20	-	-
PA3622	rpoS	sigma factor RpoS	1,20	-	-
PA1471		hypothetical protein	1,26	-	-1,15
PA5377		probable permease of ABC transporter	1,27	-	-1,22
PA4015		conserved hypothetical protein	1,30	-	-
PA0588	yeaG	conserved hypothetical protein	1,31	-	-
PA2562		hypothetical protein	1,37	-	-
PA5546		conserved hypothetical protein	1,37	-	-
PA0900		hypothetical protein	1,38	-	-
PA0129	gabP	gamma-aminobutyrate permease	1,46	-	-
PA5313	paaT	probable pyridoxal-dependent aminotransferase	1,46	-	-
PA4501	opdD	Glycine-glutamate dipeptide porin OpdP	1,65	-	-
PA0299	spuC	putrescine aminotransferase	1,69	-	-
PA5178		conserved hypothetical protein	1,69	-	-
PA5373	betB	betaine aldehyde dehydrogenase	1,72	-	-
PA0266	gabT	4-aminobutyrate aminotransferase	2,60	2,04	-1,28
PA0446		conserved hypothetical protein	3,01	-	-
PA3569	mmsB	3-hydroxyisobutyrate dehydrogenase	4,03	-	-
PA3570	mmsA	methylmalonate-semialdehyde dehydrogenase	4,30	-	-
PA0887	acsA	acetyl-coenzyme A synthetase	-	-9,61	-
PA5365	phoU	phosphate uptake regulatory protein PhoU	-	-1,20	-
PA5562	spoOJ	chromosome partitioning protein Spo0J	-	-1,19	-
PA2107		hypothetical protein	-	-1,18	-
PA3267		hypothetical protein	-	-1,16	-
PA0117		probable short chain dehydrogenase	-	-1,15	-
PA4851		hypothetical protein	-	-1,15	-
PA3841	exoS	exoenzyme S	-	-1,14	-
PA0629		conserved hypothetical protein	-	-1,13	-

PA3881		hypothetical protein	-	-1,13	-
PA4771	lldD	L-lactate dehydrogenase	-	-1,12	-
PA2064	pcoB	copper resistance protein B precursor	-	-1,12	-
PA3486		conserved hypothetical protein	-	-1,11	-
PA4548	dadA	probable D-amino acid oxidase	-	-1,11	-
PA4609	radA	DNA repair protein RadA	-	-1,11	-
PA3705	wspD	hypothetical protein	-	-1,10	-
PA1524	xdhA	xanthine dehydrogenase	-	-1,09	-
PA4955		hypothetical protein	-	-1,08	-
PA2154	ybhN	conserved hypothetical protein	-	-1,08	-
PA4183		hypothetical protein	-	-1,08	-
PA3599		probable transcriptional regulator	-	-1,07	-1,07
PA1184		probable transcriptional regulator	-	-1,07	-
PA3603	dgkA	diacylglycerol kinase	-	-1,06	-
PA1992	ercS	ErcS	-	-1,05	-
PA4581	rtcR	transcriptional regulator RtcR	-	-1,05	-
PA3067		probable transcriptional regulator	-	-1,05	-
PA2806	yqcD	conserved hypothetical protein	-	-1,04	-
PA0951	rbn	probable ribonuclease	-	-1,04	-
PA4720	trmA	tRNA (uracil-5-)-methyltransferase	-	-1,02	-
PA5135	yibQ	conserved hypothetical protein	-	-1,02	-
PA3217	cyaB	CyaB	-	-1,02	-
PA0385		hypothetical protein	-	1,02	-
PA3474	yigM	conserved hypothetical protein	-	1,03	-
PA2527	yegN	probable Resistance-Nodulation-Cell Division (RND) efflux transporter	-	1,05	-
PA3363	amiR	aliphatic amidase regulator	-	1,06	-
PA2432	bexR	bistable expression regulator, BexR	-	1,07	-
PA1290		probable transcriptional regulator	-	1,07	-
PA0916	yliG	conserved hypothetical protein	-	1,07	-
PA0566		hypothetical protein	-	1,09	-
PA1762		hypothetical protein	-	1,12	-
PA4574	yqhA	conserved hypothetical protein	-	1,13	-
PA0132	oapT	beta-alanine--pyruvate transaminase	-	1,14	-
PA2550		probable acyl-CoA dehydrogenase	-	1,18	-
PA3501		hypothetical protein	-	1,36	-
PA0745		probable enoyl-CoA hydratase/isomerase	-	1,97	-
PA5083		conserved hypothetical protein	-	4,88	-
PA5082		probable binding protein component of ABC transporter	-	5,65	-
PA0603		probable ATP-binding component of ABC transporter	-	-	-1,40
PA5528		hypothetical protein	-	-	-1,35
PA2014	liuB	methylcrotonyl-CoA carboxylase, beta-subunit	-	-	-1,26

PA4500		probable binding protein component of ABC transporter	-	-	-1,26
PA4913		probable binding protein component of ABC transporter	-	-	-1,20
PA3919	ylaK	conserved hypothetical protein	-	-	-1,20
PA0315		hypothetical protein	-	-	-1,20
PA2229	yiiM	conserved hypothetical protein	-	-	-1,19
PA3615		hypothetical protein	-	-	-1,19
PA0973	oprL	Peptidoglycan associated lipoprotein OprL precursor	-	-	-1,19
PA5153		amino acid (lysine/arginine/ornithine/histidine/octopine) ABC transporter periplasmic binding protein	-	-	-1,19
PA4198		probable AMP-binding enzyme	-	-	-1,18
PA3264	yocS	probable transporter	-	-	-1,18
PA4524	nadC	nicotinate-nucleotide pyrophosphorylase	-	-	-1,17
PA2907	cobL	precorrin-6y-dependent methyltransferase CobL	-	-	-1,17
PA1235		probable transcriptional regulator	-	-	-1,16
PA5148	yggX	conserved hypothetical protein	-	-	-1,16
PA3177		hypothetical protein	-	-	-1,15
PA4724	yadB	probable aminoacyl-transfer RNA synthetase (class I)	-	-	-1,15
ig_3265210_ 3265847	NA	NA	-	-	-1,15
PA1326	ilvA2	threonine dehydratase, biosynthetic	-	-	-1,14
PA2363		hypothetical protein	-	-	-1,14
PA4937	rnr	exoribonuclease RNase R	-	-	-1,14
PA2131	cupA4	fimbrial subunit CupA4	-	-	-1,13
PA2430		conserved hypothetical protein	-	-	-1,12
PA2251		hypothetical protein	-	-	-1,12
PA3208	ydjA	conserved hypothetical protein	-	-	-1,11
PA4419	ftsL	cell division protein FtsL	-	-	-1,11
PA4361		probable oxidoreductase	-	-	-1,11
PA1934		hypothetical protein	-	-	-1,10
PA3854		hypothetical protein	-	-	-1,10
PA5279		conserved hypothetical protein	-	-	-1,10
PA2661		hypothetical protein	-	-	-1,09
PA2495	oprN	Multidrug efflux outer membrane protein OprN precursor	-	-	-1,09
PA2447		probable transcriptional regulator	-	-	-1,08
PA3795		probable oxidoreductase	-	-	-1,07
PA3239		conserved hypothetical protein	-	-	-1,07
PA2260	kguE	hypothetical protein	-	-	-1,07
PA2440		hypothetical protein	-	-	-1,07
PA2024		probable ring-cleaving dioxygenase	-	-	-1,04
PA1714	exsD	ExsD	-	-	-1,04

PA0992	cupC1	fimbrial subunit CupC1	-	-	-1,03
PA1275	cobD	cobalamin biosynthetic protein CobD	-	-	-1,03
PA1563	ygdE	conserved hypothetical protein	-	-	1,02
PA1612		hypothetical protein	-	-	1,03
PA2912		probable ATP-binding component of ABC transporter	-	-	1,03
PA1669		hypothetical protein	-	-	1,03
PA1650		probable transporter	-	-	1,04
PA2065	pcoA	copper resistance protein A precursor	-	-	1,05
ig_1554415_ 1553675	NA	NA	-	-	1,06
PA5366	pstB	ATP-binding component of ABC phosphate transporter	-	-	1,08
PA5487		hypothetical protein	-	-	1,08
PA0022		conserved hypothetical protein	-	-	1,08
ig_2451707_ 2450765	NA	NA	-	-	1,09
PA4782		hypothetical protein	-	-	1,09
PA4319		conserved hypothetical protein	-	-	1,09
PA3254		probable ATP-binding component of ABC transporter	-	-	1,09
PA4368		hypothetical protein	-	-	1,09
PA1170		conserved hypothetical protein	-	-	1,10
PA4064		probable ATP-binding component of ABC transporter	-	-	1,10
PA5072		probableChemotaxis transducer	-	-	1,10
PA4571		probable cytochrome c	-	-	1,11
PA5189		probable transcriptional regulator	-	-	1,12
PA1300		probable sigma-70 factor, ECF subfamily	-	-	1,13
PA0907		hypothetical protein	-	-	1,13
PA3381	phnF	probable transcriptional regulator	-	-	1,13
PA3565		probable transcriptional regulator	-	-	1,14
PA0743		probable 3-hydroxyisobutyrate dehydrogenase	-	-	1,14
PA0365		hypothetical protein	-	-	1,14
PA3906		hypothetical protein	-	-	1,18
ig_4559649_ 4560294	NA	NA	-	-	1,18
PA2021		hypothetical protein	-	-	1,22

539

Generation of genetic diversity among Pseudomonas aeruginosa via horizontal gene transfer during chronic CF infections

Trine M. Markussen[1], Rasmus L. Marvig[1], Helle K. Johansen[2], Joseph S. Lam[3,] and Lars Jelsbak[1*]

[1]Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark.

[2]Department of Clinical Microbiology, University Hospital, Rigshospitalet, 2100 Copenhagen.

[3]Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario N1G2W1, Canada.

[*]Corresponding author: Lars Jelsbak, Department of Systems Biology, Technical University of Denmark, 2800 Lyngby, Denmark. Email: lj@bio.dtu.dk. Telephone: +45 45256129.

Abstract

Pseudomonas aeruginosa is the leading course of morbidity and mortality of patients suffering from cystic fibrosis. During chronic infection bacterial populations are subject to complex processes of diversification. Although *P. aeruginosa* population diversity within infected individuals has been observed, our understanding on how this diversity is generated is very poor. Knowledge about population dynamics is crucial for optimal therapeutic interventions.

Horizontal gene transfer (HGT) is an important evolutionary process by which bacteria rapidly can acquire new traits and adapt to very hostile environments. The evolutionary role and relevance of HGT in bacterial evolution within natural chronic infections remains poorly understood despite the huge interest.

To gain new insight into the evolutionary roles of HGT during chronic polyclonal infections, we examined the population structure and dynamics of a CF patient polyclonally infected with two transmissible clone types of *P. aeruginosa*.

We showed a complex population structure of the infecting clones represented by multiple co-existing sub-populations each with distinct phenotypic and genotypic signatures and identified an *in vivo* HGT of 217 kb between the two clones of *P. aeruginosa* using whole genome sequencing. The HGT was demonstrated to contribute to the observed diversity by generating novel strains with mosaic genomes that were found to persist in the CF airways.

We further showed that the HGT resulted in altered intraclonal interactions by increasing the resistance to R5 pyocin giving the recipient cell a selective advantage in bacterial warfare. The results of this study highlight the importance of HGT in shaping both individual clones and the population dynamics during chronic infection.

Introduction

Bacteria are able to react to environmental changes and rapidly evolve not only by acquisition of mutations but also by exchange of DNA. Horizontal gene transfer (HGT) has an important role in the fast bacterial evolution as HGT may give the recipient cell the ability to survive in very specific hostile environments. While fixation of beneficial point mutations can be considered a slow evolutionary process, HGT rapidly provides the recipient cell with new traits. However, not all HGT events are beneficial. In a bacterial population HGT might cause deleterious effects and thus these bacteria will be lost over time. In addition, HGT may be neutral and survival will depend on other changes. Some transfer events might confer the recipient cell a selective advantage and when this is the case, the cell lineage may be spread and maintained in the population [1]. Furthermore, HGT can lead to intraclonal diversification by incorporation of new genes or identical genes containing single nucleotide polymorphisms (SNPs) and/or indels [2]. In this way, HGT contributes to genetic diversity and sub-speciation in bacteria by blurring the boundaries between species and generating novel strains with mosaic genomes [2, 3].

It is known that bacterial diversity provides the populations with the ability to survive and persist by impair the adaptive immune response [4]. Furthermore, HGT has been highly associated with the widespread of antibiotic resistance and virulence genes [5]. This rapid acquisition of antibiotic resistance through HGT is a huge challenge in clinical settings. Consequently, HGT is exceedingly relevant for the population dynamics in natural populations including chronic infections. Despite of the huge interest, very little is known about the impact of HGT processes during chronic infections [6] and only a few studies have addressed this topic directly [3, 7]. It therefore remains of high interest to understand the evolutionary role of HGT during chronic polyclonal infection. Here we sought to investigate the role of HGT in a chronic cystic fibrosis (CF) infection.

The main reason for morbidity and mortality of patients suffering from CF is chronic infections with *Pseudomonas aeruginosa*. The opportunistic pathogen *P. aeruginosa* is ideal for studying evolution within natural chronic infections. Multiple selective pressures such as antibiotic treatment, the immune system, interactions between the infecting microbes, and the heterogeneous distribution of nutrients and chemical components in the spatially different CF environment shape the evolution and adaptation of *P. aeruginosa* [8]. Synergistic interactions between the infecting microbes as well as hostile interactions such as bacterial warfare through production and release of strain specific pyocins furthermore contribute to the evolution. These different interactions may provide the driving force for acquisition of HGT.

Two dominant and highly transmissible clones, DK1 and DK2, have been found to successfully persist and spread among Danish CF patients. Recently 55 longitudinal isolates of the DK2 clone have been sequenced in order to describe in detail the evolution dynamics of a bacterial pathogen within its human host [9]. In

this investigation a DK2 isolate sampled in 2002 from patient P24M2 (DK2-P24M2-2002) was found to have accumulated a high number of SNPs changes without being a hypermutators i.e. a strain with increased mutation rate. Previously it has been demonstrated that several hundred kilo-bases can be transferred and incorporated by homologous recombination under laboratory conditions [2]. However, to date few observations have been made on *in vivo* detection of HGT between chronic bacterial pathogens. Our main focus has therefore been to analyze the population dynamics of isolates from patient P24M2 in order to identify if the observed changes were a result of acquisition of foreign DNA and to understand the evolutionary role of HGT in polyclonal infections.

By locating the SNPs changes between DK2 reference genome and the genome of DK2-P24M2-2002 and mapping the epidemiology of the patient, we demonstrated that a large DNA fragment of 217 kb from the DK1 clone was transferred to DK2 creating a novel strain (referred to as DK1/2). The transferred DNA seemed to give the recipient cell a selective advantage showing increased resistance against R-type pyocins. Furthermore, we demonstrated that the novel strain DK1/2 persisted over time and identified among the transferred genes a potential contributor to low-level tobramycin resistance, which may have caused graduated increase of resistance.

Studying this HGT example can provide information about the bacterial responses to the different types of interactions acting on the infecting population, as well as providing the foundation for the role of HGT in shaping bacterial pathogen survival and adaptation to the human environment.

Results

Detection of recombined genomic regions in the DK2 lineage

Whole genome sequencing of isolate DK2-P24M2-2002 revealed a high number of accumulated SNPs when comparing to the reference genome of DK2 (CF333-07a; Genbank accession no. CP003149) [9]. No mutations in DNA repair associated genes were detected indicating that the high number of SNPs was not a result of increased mutation rate. Mapping of the SNP changes in isolate DK2-P24M2-2002 to the DK2 reference genome showed that 265 out of the 273 SNPs detected clustered in two regions, region 1 (251 SNPs; 81 non-synonymous and 170 synonymous) and region 2 (14 SNPs; 2 non-synonymous and 12 synonymous) and were therefore non-randomly distributed in the genome (Figure 1A). This indicated that the two regions were acquired through HGT and recombination. Region 1 covered 217 kb including the origin of replication. Region 2 was smaller and only covered 8 kb.

Recombined DNA originates from co-existing clone types in the patient

To identify a possible donor of the recombined DNA, we examined the infection history of patient P24M2 and identified a total of 30 isolates that were sampled and stored from this individual between 1987 and 2003. In this period the patient was attending the CF Centre at regular basis and *P. aeruginosa* isolates were sampled routinely. However, after 2003 samples were not available as the patient was relocated.

Using a low-resolution genotyping method (see *Materials and methods*) we determined that all 30 isolates were either DK1 or DK2 clone type (Figure 1B). The patient had been chronically infected with the DK2 clone type since 1987, and from 1992 and onward co-infected with the DK1 clone type (Figure 1B). We have previously observed polyclonal infections in which DK1 and DK2 co-exist in infected airways [10]. Since the patient relocated we were not able to investigate further microbial developments in the patient.

To explore the possibility that the co-existing DK1 clone type was the source of the recombined DNA, we sequenced the genomes of two DK1 isolates sampled from the patient in 1992 and in 2003 (isolates DK1-P24M2-1992a and DK1-P24M2-2003a) as well as a DK1 isolate sampled in 1973 from an unrelated patient (isolate DK1-P33F0-1973). We constructed a phylogenetic tree to investigate the relationship between DK2-P24M2-2002, the reference DK2 genome, and the three DK1 isolates in relation to region 1 (Figure 1C). We observed that in region 1, DK2-P24M2-2002 diverged from the reference DK2 clone with 251 SNPs as described above. In contrast, the cluster of DK1 isolates only diverged with 6 SNPs (Figure 1C). The observation that the vast majority of accumulated SNPs in isolate DK2-P24M2-2002 were highly related to DK1 clone DNA, strongly suggest that a DK2 isolate had horizontally acquired DNA from DK1 resulting in a "hybrid" isolate, referred to as DK1/2-P24M2-2002.

In the case of region 2, we observed that DK1/2-P24M2-2002 shared all 14 SNPs with the DK1 isolates. However, DK1 isolates possessed additional 19 different SNPs (Figure 1D), suggesting that region 2 had been derived from a *P. aeruginosa* clone type different from DK1. It therefore appears that DK1/2-P24M2-2002 had acquired DNA from two independent HGT events; region 1 was acquired from DK1 and the divergent region 2 was derived from an unsampled and unknown clone type. In this study, we only focused on region 1.

With this knowledge we constructed a PCR strategy for detecting the transferred regions in other isolates from this patient (see *Materials and methods*). One other hybrid DK2 strain with incorporated DK1 DNA was detected at time point 1997 (DK1/2-P24M2-1997) (Figure 1B), indicating that the horizontal transfer happened before this time point, and that the sub-lineage has been surviving in the airways of the CF patient while co-existing with both DK1 and DK2 clones.

Population structure and dynamics of patient P24M2

To further understand the population dynamics in P24M2 we measured CF-relevant phenotypes of the isolates from P24M2 (Table S1). The phenotype profiling of DK2 and DK1 isolates showed that several sub-populations existed, each with distinct functional signatures; while DK1 isolates co-existed as two sub-populations (mucoid and non-mucoid), distinct phenotypes were observed among DK2 isolates (mucoid, non-mucoid, hypermutators, and non-hypermutators) (Table S1). Sequencing of selected isolates from DK2 confirmed the co-existence of distinct sub-populations (Figure 2A). These data combined revealed a highly diverse population structure of infecting *P. aeruginosa* in patient P24M2 represented as co-existence of different sub-populations, and that this diversity was further increased by HGT and the emergence of DK1/2 (Figure 2B).

Recombination does not change the gene inventory and results in limited recombinatorial repair

Uptake and recombination of large regions of DNA could potentially introduce new genes. However, examination of the recombined region revealed that the gene inventory remained unchanged in region 1. In contrast, we found 8 indels in region 1 in addition to the 251 SNP changes identified (Table S2).

We next inspected region 1 for evidence of recombinatorial repair of genes inactivated by mutation in the DK2 lineage. We therefore sought to identify SNPs or indels in DK2 reference genome that were complemented by the transfer and recombination of DK1 DNA. We found only a total of six SNPs (but no indels) in the corresponding region in the DK2 reference genome that were complemented in DK1/2-P24M2-2002: one missense and three silent mutations in *trkA* (encoding a potassium transporter peripheral membrane component), one silent mutation in PADK2_28735 (PA5400 in PAO1), and one intergenic mutation between *soxG* and *purU2* (encoding a sarcosine oxidase gamma subunit and formyltetrahydrofolate deformylase, respectively) (Table 2). Since the DK2 lineages that successfully co-existed in parallel with DK1/2 and DK1 in the patient (Figure 2) all carried the mutated *trkA* allele it appears that ecological success of DK1/2 was not dependent on repair by recombination of this missense mutation.

The HGT changes the pyocin profile in DK1/2

With the detection of the exchange of genetic material between DK1 and DK2, we reasoned that both clone types colonized the same niche in the infected airways. Since niche establishment and protection in mixed populations is presumed to be influenced by strain-specific production of bacteriocins such as R-pyocins, we examined pyocin production in the isolates as well as their ability to grow in the presence of different R-type pyocins. R-pyocins kills rapidly and specifically closely related bacteria by binding to sugar residues of the LPS core [11-13] resulting in cell death through depolarization of the cytoplasmic membrane [14]. Five

different R-pyocins (R1-R5) has been described based on their killing abilities and the amino acid sequence of the tail fiber protein Prf15 (PA0621) [15].

R-pyocins were isolated from DK1, DK2, and DK1/2 strains and tested on a sensitive control strain (DK2-CF510-2006) to determine if the clinical strains were producing R-pyocins (Figure S1). While R1-R5 pyocins type killed WTB, it was observed that neither DK1, DK2, nor DK1/2 produced pyocins that were able to kill WTB. In support of this observation, we could not detect the tail fiber protein gene *prf15* of R1-R5 pyocins in the genomes of either of the isolates.

Next, the susceptibility of the clinical isolates was examined against the five R-type pyocins (Figure 3). It was observed that all DK1 clones were resistant to R3 and R5 where all DK2 isolates were susceptible to R5. Remarkably, the DK1/2 was significantly more resistant to R5-type pyocins compared to DK2 as approximately five fold more of DK1/2 was able to survive in the R5 spot zone (Figure 3). This suggests that the transferred DNA from DK1 had changed the DK1/2 pyocin profile making it more resistant to R5-type pyocins.

Mutation in wbpZ *affects DK1 LPS production*.

It has been demonstrated that LPS has an important role as both a protective shield and receptor site for R-pyocins [11, 13]. We therefore theorized that the increased resistance in DK1/2 was due to distinct LPS profiles and reduced availability of R-pyocin receptors. Hence we sought to identify transferred mutations in genes involved in LPS synthesis in order to suggest a molecular mechanism for the observed phenotype. Only one candidate gene (*wbpZ*) with a known role in LPS synthesis could be identified among the transferred alleles. The *wbpZ* gene encodes a glycosyltransferase involved in A-band LPS biosynthesis [16]. To test if the mutation in *wbpZ* caused increased pyocin resistance, the specific *wbpZ* allele was introduced in wild type PAO1, isolate DK1-P24M2-2003a, and DK2-P24M2-2003a isolate by allelic replacement. Furthermore, the point mutation was restored to wild type in the DK1/2 isolates. The isogenic *wbpZ* strains were then evaluated for altered pyocin resistance. Surprisingly, we observed that the mutants displayed unchanged sensitivity compared to their parental strain (data not shown), rejecting the hypothesis that the mutation in *wbpZ* caused an increase in pyocin resistance. Consequently, we examined the LPS profile of DK1, DK2, and DK1/2 in order to determine if the resistance was caused by distinct LPS profiles. Western immunoblotting using the A-band specific MAb (N1F10) showed a clear difference between DK1 and DK2 isolates. It was observed that DK1 isolates expressed A-band LPS, while DK2 seemed to have lost the A-band as no binding was observed (Figure 4B). Introducing the *wbpZ* point mutation in a DK1 isolate and reference strain PAO1 caused loss of A-band as was observed for the isolate DK1-P24M2-2003a*wbpZ* (isogenic mutant of DK1-P24M2-2003a) and PAO1wbpZ mutant. This demonstrated that *wbpZ* point

mutation caused loss of A-band LPS in PAO1 and DK1. The production of A-band was restored by complementation *in trans* using plasmid pHERD-*wbpZ* in *wbpZ* mutant strains (Figure 3C). In addition, it was observed that A-band production could not be restored by insertion of wild type *wbpZ* in DK2 isolates, indicating that DK2 may have other mutations causing loss of A-band. Both DK1 and DK2 isolates had lost B-bands (data not shown). Based on this it could be reasoned that DK1 isolates produced A-band LPS, while DK2 and DK1/2 had lost A-band production. Subsequently, the *wbpZ* mutation only had an effect in DK1 clones and was not the cause of increased pyocin resistance. Interestingly, the resistance phenotype must be caused by a genetic determinant unrelated to *wbpZ*. In relation to this, it was observed that 12% of the transferred alleles (31 SNPs and indels) were encoding membrane protein (Table S2). Of the 31 SNPs and indels, 12 candidates besides *wbpZ* were non-synonymous and may therefore represent possible candidates.

Low-level resistance.

Besides functioning as receptors for pyocins, LPS also mediates direct interaction with antibiotics. Several studies have shown that alteration of LPS structure confers increased permeability resistance to aminoglycosides [17-19]. A knockout mutant of *wbpZ* has previously been described to affect the tobramycin resistance against tobramycin making it more resistant [19]. Tobramycin competitively binds to LPS causing increased membrane permeability by disrupting the integrity of the outer membrane [19]. Therefore we performed tobramycin killing curves mimicking MIC conditions (Figure 5) in order to examine if the mutation in *wbpZ* gave rise to low level resistance.

Interestingly, the PAO1*wbpZ* mutant showed significant reduction of killing during 60 minutes by 1 µg/mL tobramycin compared to the wild type ($P < 0.05$), demonstrating that the mutation gave low-level resistance (Figure 5). This suggests that the transfer between DK1 and DK2 not only changed the pyocin sensitivity, but may also have an impact on antibiotic resistance. However, this low level resistance could not be confirmed when the *wbpZ* mutation was inserted in a DK1 or DK2 background as the clinical isolates already had increased resistance against tobramycin (Table S3).

Discussion

Our analysis of genome sequences from several DK1 and DK2 isolates provided evidence for two independent events of uptake and recombination of *P. aeruginosa* DNA into the chromosome of a DK2 isolate. One of the recombined regions (region 1) was 217 kb in size and contained 251 polymorphisms of which 245 were also found in DK1 DNA. Since DK1 and DK2 isolates co-existed in the patient, we conclude that region 1 was transferred from DK1 to DK2 generating a hybrid strain called DK1/2. The six SNPs

234 difference between region 1 of DK1/2 and the analysed DK1 isolates suggested that the specific DK1 donor
235 isolate was not among the sequenced DK1 isolates. We were unable to identify the original DK1 donor by
236 PCR screening of the remaining DK1 from P24M2, indicating that the original DK1 donor may not have been
237 sampled and stored. Alternatively, 6 SNPs accumulated in the region in DK1/2 after the recombination
238 event. The imported DNA in relation to the smaller region 2 was not related to DK1, and we suggest that
239 this region was acquired independently from a donor clonally unrelated to DK1.
240 This is a remarkable observation given the low frequency of previously detected HGT in clinical isolates,
241 indicating either that specific environment conditions existed in this patients which promoted interactions
242 and consequently transfer of genetic material between isolates, or that the increasing numbers of whole
243 genome sequencing now makes it possible to detect such events.
244 Different processes exist to transfer DNA from one organism to another. These includes transformation,
245 where naked DNA is taken up and incorporated into the chromosome; conjugation that is mediated by a
246 cell-to-cell contact and a pore where DNA can pass; and transduction, where a bacteriophage acts as the
247 vector for injection of DNA into the recipient cell [1]. It has been well documented that *P. aeruginosa* uses
248 both transduction and conjugation for acquiring DNA. However, we did not find genes related to
249 mobilization and transfer of DNA or phage related genes within a 10 kb range of the predicted boundaries
250 of the horizontally transferred regions. In contrast, *P. aeruginosa* are not known to be naturally competent
251 and no published reports of *P. aeruginosa* natural transformation is to data available. Nevertheless, the
252 data presented here indicate that transformation was the most likely mechanism to account for the
253 detected transfer of DNA. Furthermore, a number of studies have demonstrated examples of
254 recombination between different *P. aeruginosa* genotypes in environmental and clinical isolates [20-22].
255 This indicates that under certain conditions could promote natural competence and DNA transfer through
256 transformation.

257 Bacterial infection in CF is often associated with biofilm growth mode, where multiple species may coexist
258 and interact within the biofilm [23, 24]. Biofilm cells are embedded in a matrix and one of the major
259 components of the biofilm matrix is extracellular DNA released by the bacteria [25]. Transformation has
260 furthermore been demonstrated to be 10^4-fold higher in biofilms than in planktonic cells [26]. It is
261 therefore likely that the transfer of DNA in CF infections occurred in a biofilm where the two dominant
262 clones DK1 and DK2 were coexisting. Another possible mechanism involves the release and uptake of
263 membrane vesicles. During growth *P. aeruginosa* as many gram-negative bacteria release membrane
264 vesicles into the environment. These vesicles have been described to contain virulence factors from the
265 periplasm [27]. Furthermore, it has been reported that these vesicles also contain large fragments of
266 chromosomal DNA [28, 29]. Hence, it has been indicated that the vesicles may be involved in the transfer of

genetic material among similar bacterial species [30]. Transfer mediated by vesicles allows protection from DNases [30] and enhance the efficiency of DNA delivery to a recipient cell as they fuse directly to the membrane and unload the content.

The DK1/2 clone was found to be maintained during chronic CF infection (i.e. the horizontally acquired region could be traced back to an isolate sampled from 1997, and detected again in 2002). This suggests that the recipient cell may have acquired genes or mutations that conferred a selective advantage or are neutral depending on the conditions.

HGT is associated with the widespread distribution of antibiotic resistance genes, biodegradative pathways gene cluster, and pathogenicity determinants [5]. Likewise we observed that the hybrid strain had an increase resistance against the bacteriocin R5 pyocin. To our knowledge, pyocin resistance has not been described to be acquired through HGT. The increase in R5 resistance most likely would be due to loss or modification of surface receptors that would reduce R5 binding and killing of the cell.

It has been demonstrated that sugar residues of the LPS core function as receptors for R-pyocins [11, 31] although alternative receptor sites might also exist. Köhler and colleagues identified two sugar residues in the outer core that appeared to be involved in R5 pyocin recognition [11]. Furthermore, an *algC* mutant, deficient in synthesizing D-Glc of the outer membrane was found to be resistant to R5 pyocin [11]. This indicates that also mutations in genes involved in synthesis of the sugar residues of the outer core of LPS may change the sensitivity to pyocins. However, none of the enzymes involved in D-Glc synthesis were found to be mutated in the DK1/2 isolates. Instead, we identified *wbpZ*, encoding a glycosyltransferase as the only gene involved in LPS synthesis that was changed, which made it a possible candidate for R5 pyocin resistance.

The mutation did indeed cause loss of A-band LPS in a PAO1 and DK1 background, however, it did not seem to have an effect in DK2 as DK2 clones had lost A-band before introduction of the mutation. Furthermore, *wbpZ* mutants showed identical pyocin sensitivity profiles as parent strains, indicating that the increase in pyocin resistance was not a result of the mutation in *wbpZ*. Interestingly, this observation suggests that other genetic determinants caused the increase in R5 pyocin resistance. From the identified SNPs in the transferred region we observed that several membrane proteins were mutated. It is possible that the observed change in resistance may be a result of modified cell surface components working as alternative R5 pyocin receptors. Another possibility could be that the increased resistance in DK1/2 isolates was due to the combination of several mutations. As DK2 isolates had lost A-band production, the combination of multiple mutations might account for the increase pyocin resistance.

Although the mutation in *wbpZ* did not cause pyocin resistance in clinical isolates, we did observe an effect on LPS in DK1 clones, suggesting that the mutation was disrupting protein function. Modifying cell surface

component in particular LPS, which plays an important role for directly interaction with host cell receptors and antibiotics [32], has also been illustrated to affect antibiotic resistance by decreasing the aminoglycosides passage across the outer membrane [18]. Likewise, we confirmed that the mutation gave low-level tobramycin resistance in a PAO1 background but was not detectable in DK1 and DK2 backgrounds as the clinical isolates already showed increase in antibiotic resistance.

It is not known if the observed HGT was a random event or if the selective pressures acting on the bacterial population during colonization and adaptation to the CF environment triggered the transfer. The presented data here suggested that the HGT might provide selective advantages relevant in the CF environment. The transfer and recombination delivered mutations to the hybrid DK1/2 genotype that increased its resistance to strain-specific antibiotics (R5 pyocin) and protection against invading, R5 producing *P. aeruginosa* strains. In addition, the recombination provided a mutation that potentially resulted in increased drug resistance (tobramycin). Interestingly, the patient was treated with tobramycin both by inhalation and intravenously at the time of the first detection of the hybrid strain DK1/2.

The HGT may not only have an impact on the genome of the recipient cell, but also influence the population dynamics. As HGT blurs the boundaries of species by creating mixed genomes, it has been suggested that HGT is contributing to bacterial persistence by generating a more diverse population (the distributed genome hypothesis [33]). Likewise, we observed that the HGT contributed to genetic diversity and phenotypic variation in the infecting bacterial population as illustrated in the proposed model in Figure 5. This complex and highly dynamic population structure may be a way to cope with the adaptive immunity and antibiotic therapy creating a population with high diversity that provides a genetic reservoir to acquire antibiotic resistance mutations, or other virulence trait. Therefore HGT and the generated diversity may have contributed to the persistence of *P. aeruginosa* in patient P24M2.

In conclusion we observed a HGT between to dominant clones of *P. aeruginosa* infecting a CF patient. The HGT seemed to generate a more diverse population and changed the pyocin profile of the recipient cell giving the recipient cell a selective advantage. This study adds to the increasing realization of a complex and dynamic P. aeruginosa population structure in chronic CF airway infections.

Material and methods

Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are represented in Table 2. All strains were grown in Luria Bertani (LB) broth unless otherwise stated. For plasmid maintenance in *E. coli*, the medium was supplemented with 8 µg/mL tetracycline (Tc). For marker selection in *P. aeruginosa* PAO1, 50 µg/mL Tc was used.

Construction of mutants

The *wbpZ* mutant was constructed by insertion of *wbpZ*(C896T) or wild type *wbpZ* by allelic replacement. Briefly, the *wbpZ* gene was amplified from DK1/2-P24M2-2002 or PAO1 using wbpZ-fw-SacI and wbpZ-rev-XbaI (See Table S4 for primers). The fragment was digested with SacI and XbaI and ligated onto the plasmid pNJ1 creating pTM2 or pTM5 and transformed into *E. coli* CC118λpir by standard protocols [34].

The mutants were constructed by triparental matings in which the recipient received pTM2 or pTM5. *E. coli* HB101 (pRK600) was used as helper strain and *E. coli* CC118λpir as donor.

PAO1*wbpZ* was complemented by electroporation with plasmid pHERD20T containing the *wbpZ* wild type gene (pYHJL31) by the method described by Choi et al. [35]. The plasmid pYHJL31 containing WT *wbpZ* was a kind gift from Hao et al. [36].

CF patient and clinical isolates

The focus of this study was on clinical isolates of *P. aeruginosa* from patient P24M2 of genotype DK1 and DK2. The patient was attending the Danish CF Centre at Rigshospitalet in Copenhagen from 1987 to 2003 from which 30 *P. aeruginosa* isolates were collected. Samples after 2003 were not available as the patient moved hospital.

The clinical isolates and time of isolation used in this study can be observed in Table S2 together with relevant phenotypes. All the clinical isolate were genotyped in order to determine which clones were present during the infection. Genotyping was done as previous described [10]. Furthermore selected isolates (DK2-P24M2-1987, DK2-P24M2-1994, DK2-P24M2-2002a, DK2-P24M2-2003a, DK1-P24M2-1992a, DK1-P24M2-2003a, DK1/2-P24M2-1997, DK1/2-P24M2-2002, DK2-P24M2-2012) were genome sequenced in order to examine the evolutionary dynamic of this patient. The genome sequencing was performed as previously described [9]. Mutation detection, analysis and construction of the phylogenetic tree were performed as previously described [37].

Colony morphology and doubling time

Isolates were plated on PIA agar and colony morphology was inspected after 24 h of growth at 37°C using a Zeiss Axioplan 2 microscope with a 2.5/0.075× neofluar plan objective.

The doubling time was measured in LB broth. Briefly, 50 mL LB in 250 mL shaking flask with baffles were inoculating an OD_{600nm} of 0.01. The flask was incubated at 37°C with shaking at 240 rpm in a New Brunswick Scientific incubator. Growth rates were measured by monitoring the OD_{600nm} during growth. The

365 measurements were done at least in duplicates. Growth rates were calculated from the exponential part of
366 the growth curves and expressed as doubling times in minutes (ln(2)/minutes).
367
368 *PCR strategy for detecting HGT*
369 A PCR strategy was constructed for detecting if the DNA in region 1 could be classified as DK1 or DK2 DNA
370 of isolates from patient P24M2. Primers were constructed for amplification of two independent fragments
371 of region 1. Primers are listed in Table 3. After amplification, the PCR product was digested with either KpnI
372 or EcoRI to detect if the fragment was identical to DK1 or DK2 DNA. An isolate with DK1 DNA in region 1
373 would only be digested with KpnI, where a DNA fragment that originated from DK2 only would be cut by
374 EcoRI. In this way we could detect if a DK2 clone had DK1 DNA in region 1.
375
376 *Pyocin susceptibility test*
377 The pyocin susceptibility of the clinical isolates was determined as previously described [11]. Briefly, strains
378 were grown overnight (ON) at 37°C in LB broth. Cultures were spun down at 11,000xg for 5 min and 5%
379 chloroform was added to the supernatant to kill the residual bacteria. This solution was used as a source of
380 pyocins. The indicator strains were grown ON at diluted to an OD_{600} = 0.64 in 0.9% NaCl and 3 μL were
381 added to 3 mL of top agar (0.4% LB agar) and poured onto a LB medium plate. 4 μL of the pyocin solution
382 was spotted on the agar plate and incubated ON at 30°C. Clear lysis zones were considered to be the result
383 of pyocin activity.
384 R-pyocins were isolated from the following strains PAK (R1-pyocin producer), PAO1 (R2-pyocin producer),
385 NIH-I (R3-pyocin producer), NIH-H (R4-pyocin producer) and NIH-1 (R5-pyocin). The clinical isolate WTB was
386 used as control as it was sensitive to all R-type pyocins.
387 Fluorescence-tagged strains were constructed as previously described with green fluorescent protein (GFP)
388 or monomeric red fluorescent protein (mRFP1) using miniTn7 constructs [38]. The tagged strains were
389 tested for their pyocin susceptibility as described above and the differences between DK1/2 and DK2 was
390 visualized by inspecting the fluorescence of the strains in the R5 undiluted spot zone with a Zeiss Axioplan 2
391 microscope using a 2.5/0.075× neofluar plan objective.
392
393 *LPS purification and analysis*
394 LPS was prepared using the proteinase K digestion method described by Hitchcock and Brown [1]. LPS was
395 13nalysed by electrophoresis on 12% SDS-polyacrylamide gels. Banding patterns were visualized by the
396 silver staining method of Fomsgaard *et al*. [30]. Western immunoblot analysis was performed using Mab
397 N1F10 (specific for A-band LPS [39]) [40]. The secondary antibody alkaline phosphatase-conjugated goat

anti-mouse Fab_2 (Jackson ImmunoResearch) was used. Finally the blots were developed using nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolylphosphate (BCIP) as described previously [40]. In order to standardize the samples, the lower part of the LPS gel was silver stained, while the upper part was blotted with N1F10.

Tobramycin killing curves

The time kill curves of PAO1 and mutant PAO1*wbpZ* were performed in triplicates in microtitre plates as described by Schurek *et al.* [19] with modifications. Briefly, isogenic mutants were grown in cation-adjusted MHB (CAMHB) at 37°C ON. Cultures were inoculated to an optical density at 600 nm of 0.01 and grown to OD_{600} of 0.5. Cultures were inoculated in microtitre plates containing 2XMIC of tobramycin (1 µg/mL for PAO1) at a final concentration of OD_{600} 0.005. Separate wells were inoculated for each time point. Aliquots were plated on LB plates at 0, 10, 20, 30, 40, 50, and 60 minutes and incubated ON at 37°C for counting.

Acknowledgement

We would like to thank the Lab of J. Lam (Department of Molecular and Cellular Biology, University of Guelph) in particular D. Kocincova and Y. Hao for assistance with LPS purification and analysis. We would also like to thank A. J. Christensen for construction of the GFP tagged strains and the Infection Microbiology Group for helpful discussions.

Author contributions: T.M.M., R.L.M. and L.J. conceived the study and designed research; T.M.M. and R.L.M. performed research; H.K.J and J.S.L. contributed with materials, reagents and analysis tools; T.M.M., R.L.M, and L.J., analysed the data and T.M.M. and L.J. wrote the manuscript.

References

1. Thomas, C.M. and K.M. Nielsen, *Mechanisms of, and barriers to, horizontal gene transfer between bacteria.* Nat Rev Microbiol, 2005. **3**(9): p. 711-21.
2. Schurmans, S., et al., *Anti-LFA-1 (CD11a) monoclonal antibody interferes with neonatal induction of tolerance to alloantigens.* European journal of immunology, 1994. **24**(4): p. 985-90.
3. Arsenault, A.L., B.W. Frankland, and F.P. Ottensmeyer, *Vectorial sequence of mineralization in the turkey leg tendon determined by electron microscopic imaging.* Calcif Tissue Int, 1991. **48**(1): p. 46-55.
4. Ehrlich, G.D., et al., *Bacterial plurality as a general mechanism driving persistence in chronic infections.* Clinical orthopaedics and related research, 2005(437): p. 20-4.
5. de la Cruz, F. and J. Davies, *Horizontal gene transfer and the origin of species: lessons from bacteria.* Trends in microbiology, 2000. **8**(3): p. 128-33.
6. Hiller, N.L., et al., *Generation of genic diversity among Streptococcus pneumoniae strains via horizontal gene transfer during a chronic polyclonal pediatric infection.* PLoS Pathog, 2010. **6**(9): p. e1001108.

7. Bilodeau, M., et al., *EMG power spectra of elbow extensors during ramp and step isometric contractions.* European journal of applied physiology and occupational physiology, 1991. **63**(1): p. 24-8.
8. Yang, L., L. Jelsbak, and S. Molin, *Microbial ecology and adaptation in cystic fibrosis airways.* Environ Microbiol, 2011. **13**(7): p. 1682-9.
9. Marvig, R.L., et al., *Genome Analysis of a Transmissible Lineage of Pseudomonas aeruginosa Reveals Pathoadaptive Mutations and Distinct Evolutionary Paths of Hypermutators.* PLoS genetics, 2013. **9**(9): p. e1003741.
10. Jelsbak, L., et al., *Molecular epidemiology and dynamics of Pseudomonas aeruginosa populations in lungs of cystic fibrosis patients.* Infect Immun, 2007. **75**(5): p. 2214-24.
11. Kohler, T., V. Donner, and C. van Delden, *Lipopolysaccharide as shield and receptor for R-pyocin-mediated killing in Pseudomonas aeruginosa.* J Bacteriol, 2010. **192**(7): p. 1921-8.
12. Rivera, M., et al., *Heterogeneity of lipopolysaccharides from Pseudomonas aeruginosa: analysis of lipopolysaccharide chain length.* J Bacteriol, 1988. **170**(2): p. 512-21.
13. Campos, J.L., et al., *[Diagnosis by DNA analysis in familial isolated growth hormone deficiency type I-A].* Medicina clinica, 1989. **92**(7): p. 261-3.
14. Rivera, M. and E.J. McGroarty, *Analysis of a common-antigen lipopolysaccharide from Pseudomonas aeruginosa.* J Bacteriol, 1989. **171**(4): p. 2244-8.
15. Bilodeau, L., et al., *Technetium-99m sestamibi tomography in patients with spontaneous chest pain: correlations with clinical, electrocardiographic and angiographic findings.* Journal of the American College of Cardiology, 1991. **18**(7): p. 1684-91.
16. Rocchetta, H.L., et al., *Three rhamnosyltransferases responsible for assembly of the A-band D-rhamnan polysaccharide in Pseudomonas aeruginosa: a fourth transferase, WbpL, is required for the initiation of both A-band and B-band lipopolysaccharide synthesis.* Mol Microbiol, 1998. **28**(6): p. 1103-19.
17. Bryan, L.E., K. O'Hara, and S. Wong, *Lipopolysaccharide changes in impermeability-type aminoglycoside resistance in Pseudomonas aeruginosa.* Antimicrob Agents Chemother, 1984. **26**(2): p. 250-5.
18. Kadurugamuwa, J.L., J.S. Lam, and T.J. Beveridge, *Interaction of gentamicin with the A band and B band lipopolysaccharides of Pseudomonas aeruginosa and its possible lethal effect.* Antimicrob Agents Chemother, 1993. **37**(4): p. 715-21.
19. Schurek, K.N., et al., *Novel genetic determinants of low-level aminoglycoside resistance in Pseudomonas aeruginosa.* Antimicrob Agents Chemother, 2008. **52**(12): p. 4213-9.
20. Chung, J.C., et al., *Genomic variation among contemporary Pseudomonas aeruginosa isolates from chronically infected cystic fibrosis patients.* J Bacteriol, 2012. **194**(18): p. 4857-66.
21. Kiewitz, C. and B. Tummler, *Sequence diversity of Pseudomonas aeruginosa: impact on population structure and genome evolution.* J Bacteriol, 2000. **182**(11): p. 3125-35.
22. Wiehlmann, L., et al., *Population structure of Pseudomonas aeruginosa.* Proc Natl Acad Sci U S A, 2007. **104**(19): p. 8101-6.
23. Eberl, L. and B. Tummler, *Pseudomonas aeruginosa and Burkholderia cepacia in cystic fibrosis: genome evolution, interactions and adaptation.* International journal of medical microbiology : IJMM, 2004. **294**(2-3): p. 123-31.
24. Yang, L., et al., *Current understanding of multi-species biofilms.* International journal of oral science, 2011. **3**(2): p. 74-81.
25. Whitchurch, C.B., et al., *Extracellular DNA required for bacterial biofilm formation.* Science, 2002. **295**(5559): p. 1487.
26. Bernier, R. and H.H. Arsenault, *Deconvolution of two-dimensional images with zeros in the transfer function.* Applied optics, 1991. **30**(35): p. 5163-8.
27. Kolling, G.L. and K.R. Matthews, *Export of virulence genes and Shiga toxin by membrane vesicles of Escherichia coli O157:H7.* Appl Environ Microbiol, 1999. **65**(5): p. 1843-8.

28. Kadurugamuwa, J.L. and T.J. Beveridge, *Virulence factors are released from Pseudomonas aeruginosa in association with membrane vesicles during normal growth and exposure to gentamicin: a novel mechanism of enzyme secretion.* J Bacteriol, 1995. **177**(14): p. 3998-4008.
29. Renelli, M., et al., *DNA-containing membrane vesicles of Pseudomonas aeruginosa PAO1 and their genetic transformation potential.* Microbiology, 2004. **150**(Pt 7): p. 2161-9.
30. Fomsgaard, A., M.A. Freudenberg, and C. Galanos, *Modification of the silver staining technique to detect lipopolysaccharide in polyacrylamide gels.* J Clin Microbiol, 1990. **28**(12): p. 2627-31.
31. Meadow, P.M. and P.L. Wells, *Receptor sites for R-type pyocins and bacteriophage E79 in the core part of the lipopolysaccharide of Pseudomonas aeruginosa PAC1.* J Gen Microbiol, 1978. **108**: p. 339-343.
32. King, J.D., et al., *Review: Lipopolysaccharide biosynthesis in Pseudomonas aeruginosa.* Innate Immun, 2009. **15**(5): p. 261-312.
33. Ehrlich, G.D., et al., *The distributed genome hypothesis as a rubric for understanding evolution in situ during chronic bacterial biofilm infectious processes.* FEMS immunology and medical microbiology, 2010. **59**(3): p. 269-79.
34. Finch, L., H. Barbeau, and B. Arsenault, *Influence of body weight support on normal human gait: development of a gait retraining strategy.* Physical therapy, 1991. **71**(11): p. 842-55; discussion 855-6.
35. Choi, K.H., A. Kumar, and H.P. Schweizer, *A 10-min method for preparation of highly electrocompetent Pseudomonas aeruginosa cells: application for DNA fragment transfer between chromosomes and plasmid transformation.* Journal of microbiological methods, 2006. **64**(3): p. 391-7.
36. Hao, Y., et al., *Five new genes are important for common polysaccharide antigen biosynthesis in Pseudomonas aeruginosa.* mBio, 2013. **4**(1): p. e00631-12.
37. Yang, L., et al., *Evolutionary dynamics of bacteria in a human host environment.* Proc Natl Acad Sci U S A, 2011. **108**(18): p. 7481-6.
38. Koch, B., L.E. Jensen, and O. Nybroe, *A panel of Tn7-based vectors for insertion of the gfp marker gene or for delivery of cloned DNA into Gram-negative bacteria at a neutral chromosomal site.* Journal of microbiological methods, 2001. **45**(3): p. 187-95.
39. Lam, M.Y., et al., *Occurrence of a common lipopolysaccharide antigen in standard and clinical strains of Pseudomonas aeruginosa.* J Clin Microbiol, 1989. **27**(5): p. 962-7.
40. de Kievit, T.R. and J.S. Lam, *Monoclonal antibodies that distinguish inner core, outer core, and lipid A regions of Pseudomonas aeruginosa lipopolysaccharide.* J Bacteriol, 1994. **176**(23): p. 7129-39.
41. Herrero, M., V. de Lorenzo, and K.N. Timmis, *Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria.* J Bacteriol, 1990. **172**(11): p. 6557-67.
42. Kessler, B., V. de Lorenzo, and K.N. Timmis, *A general system to integrate lacZ fusions into the chromosomes of gram-negative eubacteria: regulation of the Pm promoter of the TOL plasmid studied with all controlling elements in monocopy.* Mol Gen Genet, 1992. **233**(1-2): p. 293-301.
43. Holloway, B.W., V. Krishnapillai, and A.F. Morgan, *Chromosomal genetics of Pseudomonas.* Microbiol Rev, 1979. **43**(1): p. 73-102.
44. Arsenault, A.L., et al., *Apatite determination and distribution in an early stage of gallbladder calcification.* Ultrastructural pathology, 1991. **15**(2): p. 175-83.
45. Lambertsen, L., C. Sternberg, and S. Molin, *Mini-Tn7 transposons for site-specific tagging of bacteria with fluorescent proteins.* Environ Microbiol, 2004. **6**(7): p. 726-32.
46. Skolimowski, M., et al., *Modular microfluidic system as a model of cystic fibrosis airways.* Biomicrofluidics, 2012. **6**(3): p. 34109.

Figures:

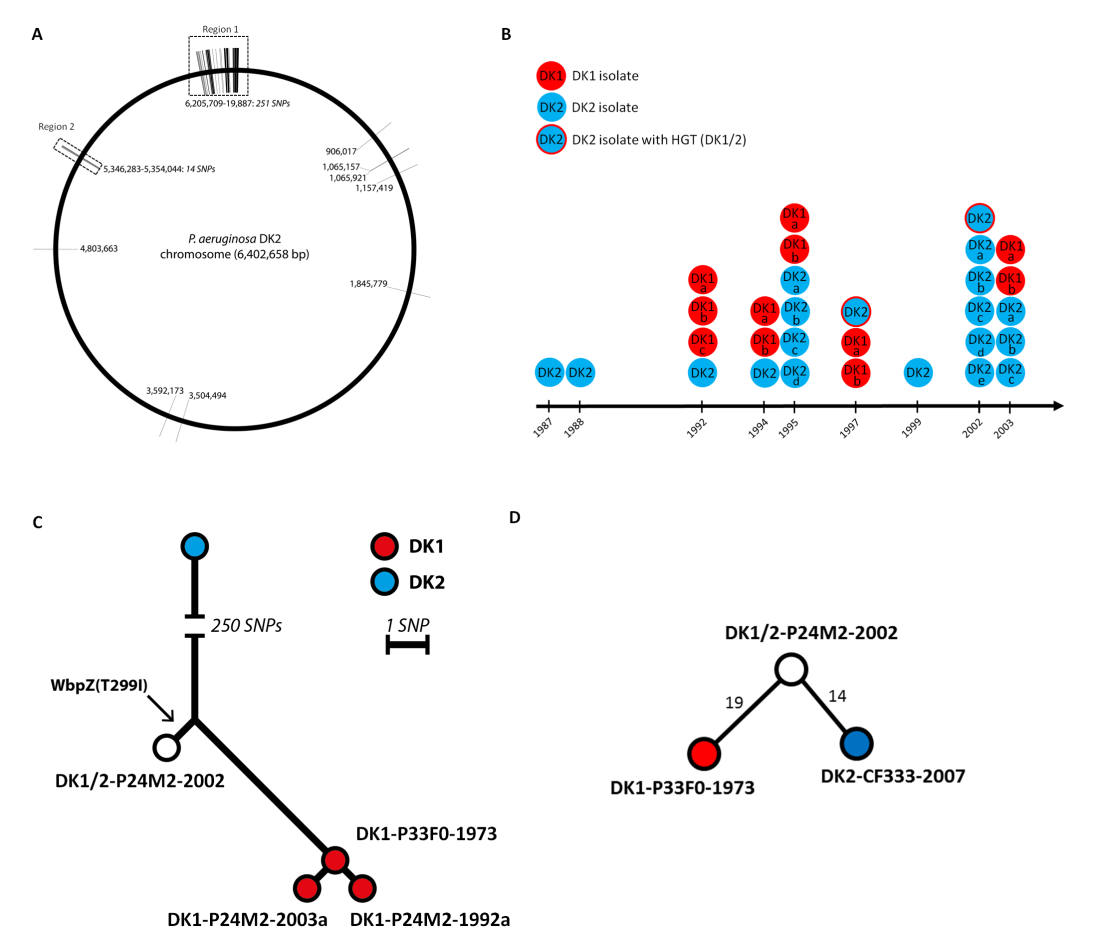


Figure 1. Detection of horizontal gene transfer between DK2 and DK1. A) SNPs of P24M2-02 mapped against *P. aeruginosa* DK2 chromosome. The SNPs cluster mainly in two regions; region 1 (position 6,205,709 – 19,887) and region 2 (position 5,346,283 – 5,354,044). Black lines indicate the SNP difference of DK2-P24M2-2002 compared with CF333-07. B) Overview of *P. aeruginosa* clones isolated from patient P24M2. Sampling time and clone genotype are showed; red dots represent DK1 isolates while blue dots represent DK2. Blue dots with red circles represent DK1/2 isolates. The isolates are named by their clone type, and by the patient from whom they were isolated and their isolation year (e.g. DK2-P24M2-1987). C) Phylogenetic tree of Region 1. DK2-P24M2-2002 and the DK1 cluster are separated with 6 SNPs, while DK2-P24M2-2002 and DK2 are separated with 251 SNPs. The 6 SNPs differences were 4 silent mutations and to missense found in *wbpZ* and PADK2_28940. The phylogenetic tree is based on the 251 identified SNPs of region 1 from whole-genome sequencing and lengths of branches are proportional to the SNPs between the isolates. D) Representation of differences between DK1, DK2, and DK1/2 in region 2. DK1/2-P24M2-2002 and DK2-CF333-2007 are separated by 14 SNPs, while DK1/2-P24M2-2002 and DK1-CF33-1973 is separated with 19 SNPs.

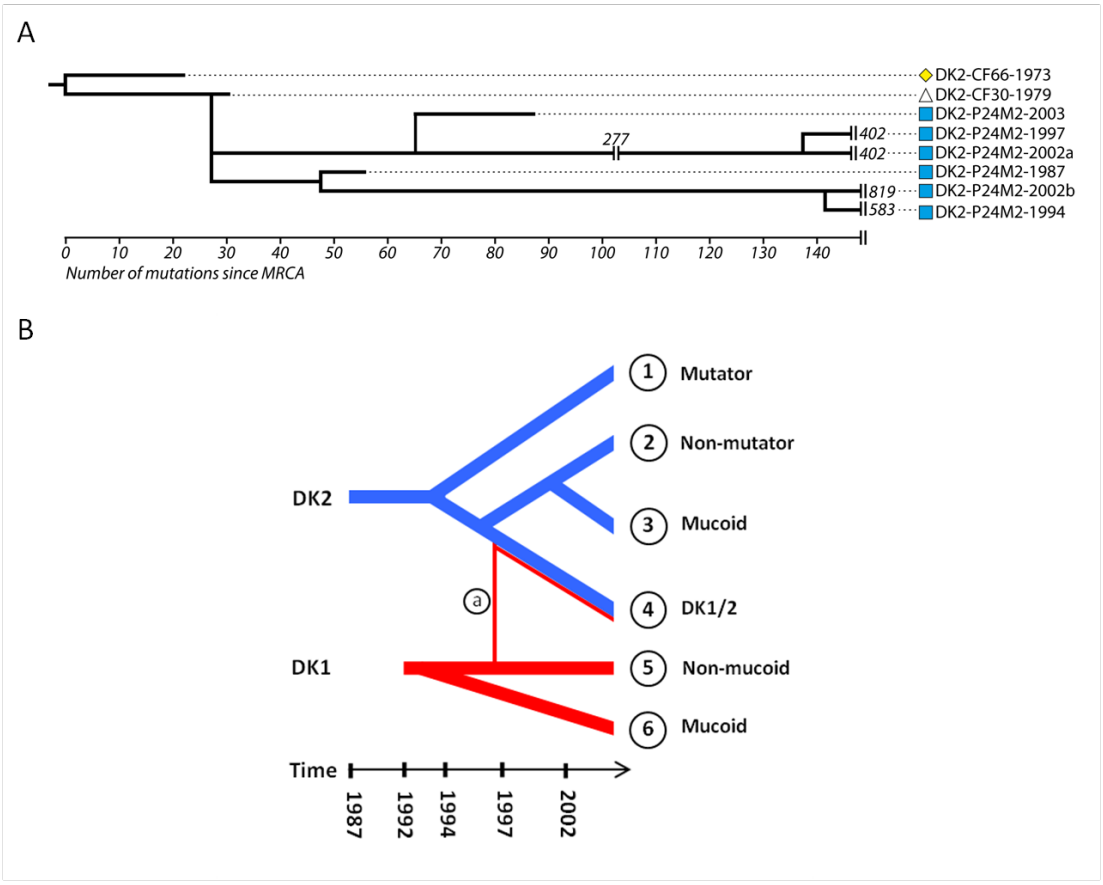


Figure 2. Population structure of P24M2. A) Phylogenetic three of DK2 and DK1/2 isolates from P24M2. The three is based on SNPs identified by whole genome sequencing and lengths of branches are proportional to the number of mutations. Isolate DK2-CF66-1973 and DK2-CF30-1979 (from [9]) was used as an outgroup to determine the root of the P24M2 tree. B) Model of population structure and diversity of five different sub-populations (1-5) from patient P24M2. (1) Sub-population of DK2 found to be mutator. (2) Non-mutators found to be closely related to DK1/2 (4), that at a given time point before 1997 gained horizontally transferred region (a) from DK1. (3) Sub-population of DK2 occurring in 2002 that was found to be mucoid. DK1 showed co-existence of two sub-populations: (5) was a non-mucoid sub-population and (6) was found to be mucoid. The model is made based on both phenotypic and genotypic data.

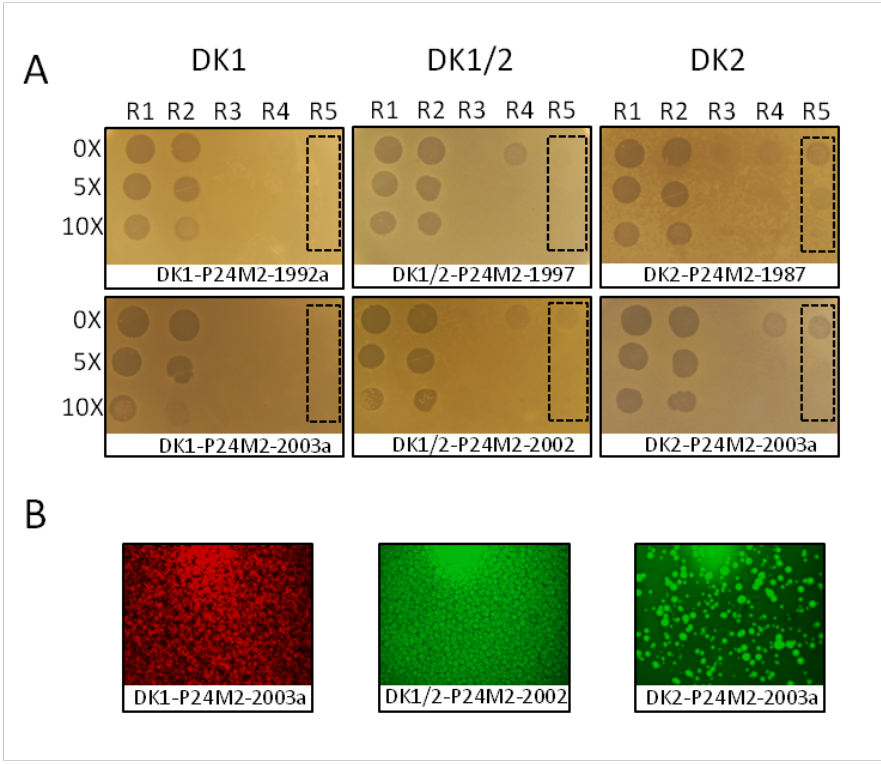


Figure 3. Pyocin killing assay. A) The killing ability of R1-R5 pyocins isolated from *P. aeruginosa* strains against DK1, DK2 and DK1/2 isolates were measured by spotting R1 (PAK), R2 (PAO1), R3, (NIH-I), R4 (NIH-H), and R5 (NIH-1) on top of the bacterial lawn in different concentrations. B) Spot zones of R5 pyocin (undiluted) on lawns of fluorescently tagged strains with either GFP (DK1/2-P24M2-2002 and DK2-P24M2-2003a) or mRFP1 (DK1-P24M2-2003a).

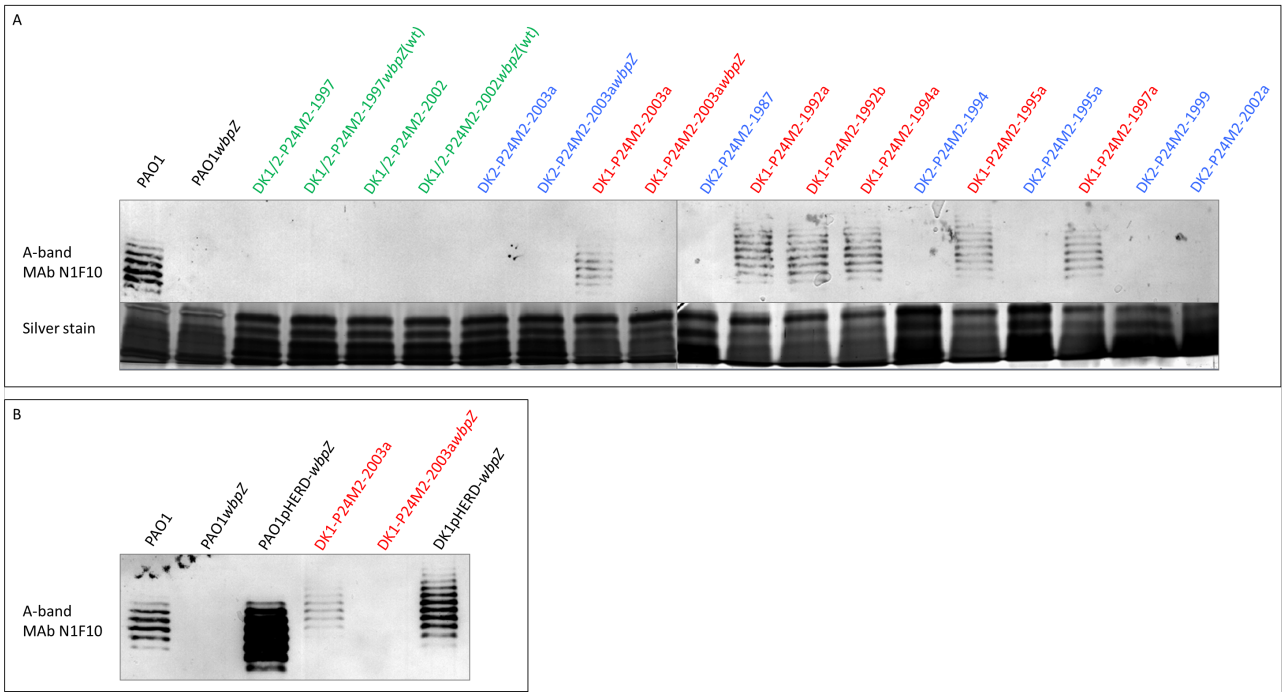


Figure 4: SDS-polyacrylamide gel and western immunoblot analysis of LPS from PAO1, *wbpZ*, and clinical isolates from P24M2. A) (Top) Western immunoblot obtained with Mab N1F10 (A-band specific); (bottom) silver stained SDS-PAGE obtained from the same gel. B) Western immunoblot obtained with Mab N1F10. Lanes 1 to 20 are identical in subfigure A and B. Colours indicate the genotype DK1 (red), DK2 (blue), and DK1/2 (green) of the clinical isolates. Note that the differences of reactivity of LPS observed between samples are not due to loading variation. To ensure that the level of reactivity of LPS with Mab N1F10 was constant among the different clinical isolates and mutants, sample loads were simultaneously monitored on silver stain (lower part of SDS-PAGE) and Western blot by reacting the blots with Mab N1F10.

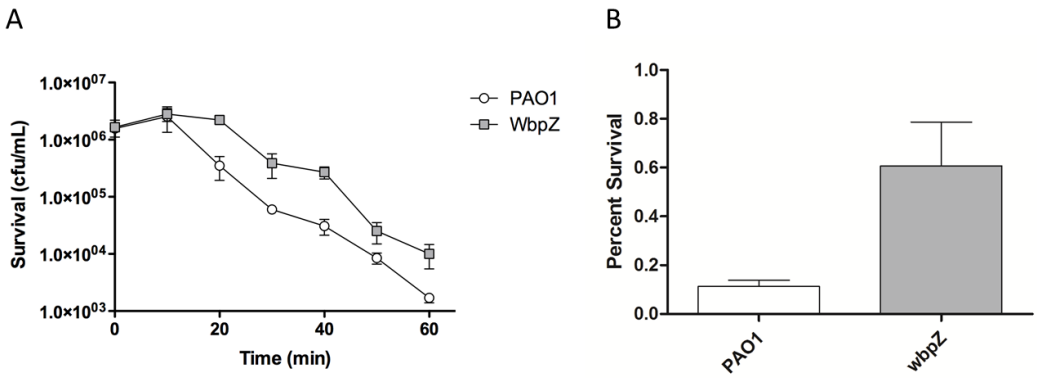


Figure 5. Tobramycin killing curves. A) Impact of *wbpZ* point mutation on tobramycin killing of strain PAO1. Error bars represent the standard deviations calculated from at least triplicates. B) Percent survival of PAO1 and PAO1*wbpZ* point mutant after 60 minutes of exposure to tobramycin. Error bars represent the standard deviations calculated from at least triplicates.

Tables:

Table 1. Complementation of DK1/2 by HGT from DK1. Nucleotides for PAO1, DK1, DK2 and DK1/2 is shown for the complemented position.

Genome position	Type of mutation	Locus tag	Gene name	PAO1	DK1	DK2	DK1/2	PseudoCAP Function Class
18494	Missense	PADK2_00080	*trkA*	G	G	A	G	potassium transporter peripheral membrane component
17391	Silent	PADK2_00080	*trkA*	A	A	C	A	potassium transporter peripheral membrane component
17727	Silent	PADK2_00080	*trkA*	G	G	T	G	potassium transporter peripheral membrane component
17751	Silent	PADK2_00080	*trkA*	A	A	G	A	potassium transporter peripheral membrane component
6217415	Silent	PADK2_28735		T	T	G	T	electron transfer flavoprotein, alpha subunit
6237772	Intergenic	PADK2_28830//PADK2_28835		G	G	A	G	27 downstream sarcosine oxidase gamma subunit//23 upstream formyltetrahydrofolate deformylase

Table 2. Bacterial strains and plasmids used in this study.

Strains, plasmids or primers	Relevant genotype or phenotype	Source of reference
Strains		
CC118λpir	*E.coli* CC18 lysogenized with λ-pir phage	[41]
HB101	*E. coli, recA,* thi, pro, leu, hsd RM^{+}; Sm^{R}	[42]

PAO1	*P. aeruginosa* Wild-type	[43]
PAO1*wbpZ*	PAO1 *wbpZ* (C896T)	This study
CF333-07	DK2 isolate from patient CF333 from 2007, GenBank accession no. CP003149)	[37]
DK1/2-P24M2-1997	DK2 isolate from patient P24M2 sampled in 1997 containing the HGT.	This study
DK1/2-P24M2-2002	DK2 isolate from patient P24M2 sampled in 2002 containing the HGT.	This study
DK1-P24M2-2003a	DK1 isolate from patient P24M2 sampled in 2003	This study
DK2-P24M2-2003a	DK2 isolate from patient P24M2 sampled in 2003	This study
DK1/2-P24M2-1997*wbpZ*(wt)	DK1/2-P24M2-1997 with wt *wbpZ*	This study
DK1/2-P24M2-2002*wbpZ*(wt)	DK1/2-P24M2-2002 with wt *wbpZ*	This study
DK2-P24M2-2003a*wbpZ*	DK2-P24M2-2003a with *wbpZ*(C896T)	This study
DK1-P24M2-2003a*wbpZ*	DK1-P24M2-2003a with *wbpZ*(C896T)	This study
PAO1pHERD-*wbpZ*	PAO1*wbpZ* complemented *in trans* with pYHJL31	This study
DK1pHERD-*wbpZ*	DK1-P24M2-2003a*wbpZ* complemented *in trans* with pYHJL31	This study
Plasmids		
pNJ1	Tc^R, $sacB^+$ R6K *ori*, mobRP4, allelic replacement vector derived from pDS132	[44]
pTM2	, pNJ1 with SacI-XbaI fragment containing PAO1 *wbpZ*(C896T)	This study
pTM5	, pNJ1 with *SacI-XbaI* fragment containing PAO1 wt *wpbZ*	This study
pYHJL31	,pHERD20T with *SacI-XbaI* fragment containing PAO1 wt *wbpZ*	[36]
pTMM	, pNJ1 with SacI-XbaI fragment containing CF333-07 PADK2_24865	This study
pTMI	, pNJ1 with SacI-XbaI fragment containing CF333-07 PADK2_24875//PADK2_24880	This study
pRK600	Cm^R, *ori* ColE1 RK2-Mob^+ RK2-Tra^+ helper plasmid in triparental matings	[42]
miniTn7 GFP	miniTn7(GM)$P_{A1/04/03}$::gfp Gm^R	[45]
miniTn7 mRFP1	miniTn7(GM)$P_{A1/04/03}$::mRFP1 Gm^R	[46]
pUX-BF13	R6K replicon-based helper plasmid, providing the Tn7 transposition function in trans. Amp^+, mob^+	[38]

Table S2. List of Single Nucleotide Polymorphims and indels transferred from DK1 to DK1/2 in region 1.

Table S2 is available here: **https://www.dropbox.com/sh/qqbt2ezwlhl6ahl/1EsaTipuGr**

Table S3. Primers used for detecting HGT.

Primers	Enzyme	Cuts genotype	Oligonucleotide sequence (5'-3')	Size of PCR product(s) (bp)	Size of fragments when digested
KpnI_fw	KpnI	DK1	TCCAGGTACGTGGCGAAG	834	393, 441
KpnI_rev			CGTGGAAACCATCGAACAG		
EcoRI_fw	EcoRI	DK2	GGTAGTGCGGATCAACCTGT	968	425, 543
EcoRI_rev			GCTCCGGGTGTTTGACATAG		

Supplementary information

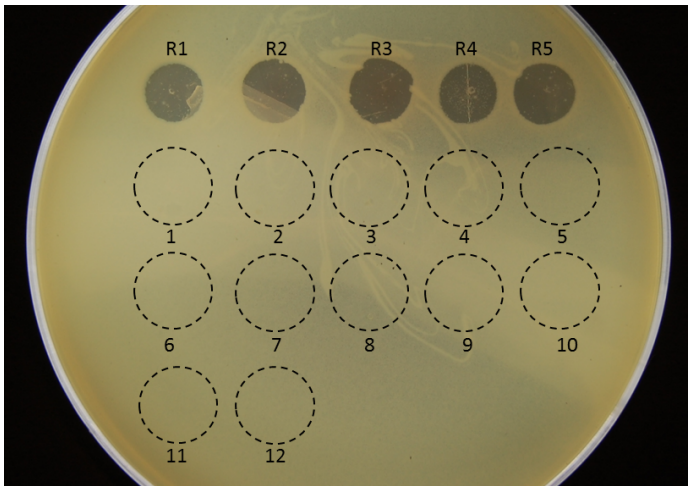


Figure S1. Pyocin susceptibility test. The pyocin isolated from DK2-P24M2-1987 (1), DK1-P24M2-1992a (2), DK1-P24M2-1994a (3), DK1-P24M2-1995a (4), DK2-P24M2-1995a (5), DK1-P24M2-1997a (6), DK1/2-P24M2-1997 (7), DK2-P24M2-1999 (8), DK2-P24M2-2002a (9), DK1/2-P24M2-2002 (10), DK1-P24M2-2003s (11), and DK2-P24M2-2003a (12) were tested against the sensitive control stain DK2-CF510-2006 together with the five known R-type pyocins (R1-R5).

Table S1. Phenotypic characterization of isolates from patient P24M2.

name	muc/non[a]	Genotype	Region 1	Doubling time (min)	Morphology[b]	Hypermutator[c]
DK1-P24M2-1992a	nm	DK1	DK1	45.5	green	ND
DK1-P24M2-1992b	m	DK1	DK1	64.82	Brown	-
DK1_P24M2-1992c	m	DK1	DK1	56.7	green	ND
DK1-P24M2-1994a	nm	DK1	DK1	50.5	green	ND
DK1-P24M2-1994b	nm	DK1	DK1	69.67	Yellow	ND
DK1-P24M2-1995b	m	DK1	DK1	34.7	green	ND
DK1-P24M2-1995a	m	DK1	DK1	68.7	green	ND
DK1-P24M2-1997b	m	DK1	DK1	61.51	Brown	ND
DK1-P24M2-1997a	nm	DK1	DK1	66.97	Green	ND
DK1-P24M2-2003b	m	DK1	DK1	60.83	Brown	-
DK1-P24M2-2003a	nm	DK1	DK1	62.04	Green	ND
DK2-P24M2-1987	nm	DK2	DK2	101.85	Colorless, sheen	-
DK2-P24M2-1988	nm	DK2	DK2	110.31	Colorless, sheen	ND
DK2-P24M2-1992	nm	DK2	DK2	109.7	White/colorless, sheen	ND
DK2-P24M2-1994	nm	DK2	DK2	77.6	Yellow	+
DK2-P24M2-1995a	nm	DK2	DK2	89.61	Colorless, sheen	ND
DK2-P24M2-1995b	nm	DK2	DK2	90.4	Colorless, sheen	ND
DK2-P24M2-1995c	nm	DK2	DK2	84.6	Yellow/sheen	ND
DK2-P24M2-1995d	nm	DK2	DK2	80.5	Yellow/sheen	ND
DK1/2-P24M2-1997	nm	DK2	DK1	67.35	Colorless, sheen	-
DK2-P24M2-1999	nm	DK2	DK2	89.4	Yellow/sheen	ND
DK1/2-P24M2-2002	nm	DK2	DK1	76.61	Colorless, sheen	-
DK2-P24M2-2002a	nm	DK2	DK2	54.74	White/colorless	ND
DK2-P24M2-2002b	m	DK2	DK2	80.59	White/colorless	+
DK2-P24M2-2002c	m	DK2	DK2	67.2	White/colorless	ND
DK2-P24M2-2002d	nm	DK2	DK2	48.5	White/colorless	ND
DK2-P24M2-2002e	nm	DK2	DK2	ND	White/colorless	ND
DK2-P24M2-2003a	nm	DK2	DK2	96.95	White/colorless	-
DK2-P24M2-2003b	nm	DK2	DK2	80.74	White/colorless	ND
DK2-P24M2-2003c	m	DK2	DK2	111.83	White/colorless	-

[a] muc/non; m, mucoid; nm, non-mucoid

[b] Color of colonies grown on LB agar plates. Sheen; visible metallic irident sheen when grown on LB agar.

[c] Hypermutation was determined based on sequencing data showing elavated mutation rate and/or mutations in DNA repair associated genes

Table S3. Tobramycin MIC values of DK1, DK2, and DK1/2 strains

Strain	Tobramycin MIC (μg/mL)
PAO1	0.25
DK1/2-P24M2-1997	16
DK1/2-P24M2-1997wbpZ(wt)	16
DK1/2-P24M2-2002	8
DK1/2-P24M2-2002wbpZ(wt)	8
DK1-P24M2-2003a	0.5
DK1-P24M2-2003awbpZ	0.5
DK2-P24M2-2003a	4
DK2-P24M2-2003awbpZ	4

Table S4. Primers used in this study

Primers	Oligonucleotide sequence (5'-3')	Reference
wbpZ-fw-SacI	GAGAGAGCTCGCATCCAGAGCCGTCGCCAG	This study
wbpZ-rev-XbaI	ACATTTAGATCGTCGGGGTGATGCGCTAC	This study
M_fw_SacI	GAGAGAGCTCACTTCGGTCAGGCCTTCAAT	This study
M_rev_XbaI	ACATTTAGACTGCACGTCCGACAGTTG	This study
I_fw_SacI	GAGAGAGCTCATAGACAGGCCAGGGGACTT	This study
I_rev_XbaI	ACATTTAGAAAAAAGCGCCGATTGTACTG	This study