# Evolutionary dynamics of bacteria in a human host environment

Lei Yang[a,1], Lars Jelsbak[a,1], Rasmus Lykke Marvig[a], Søren Damkiær[a], Christopher T. Workman[a], Martin Holm Rau[a], Susse Kirkelund Hansen[a], Anders Folkesson[a], Helle Krogh Johansen[b], Oana Ciofu[c], Niels Høiby[b,c], Morten O. A. Sommer[a,d], and Søren Molin[a,2]

[a]Department of Systems Biology, Technical University of Denmark, 2800 Lyngby, Denmark; [b]Department of Clinical Microbiology, Rigshospitalet, University Hospital, 2100 Copenhagen, Denmark; [c]Institute for International Health, Immunology and Microbiology, University of Copenhagen, 2200 Copenhagen, Denmark; and [d]Department of Genetics, Harvard Medical School, Boston, MA 02115

Edited* by Richard E. Lenski, Michigan State University, East Lansing, MI, and approved March 25, 2011 (received for review December 7, 2010)

**Laboratory evolution experiments have led to important findings relating organism adaptation and genomic evolution. However, continuous monitoring of long-term evolution has been lacking for natural systems, limiting our understanding of these processes in situ. Here we characterize the evolutionary dynamics of a lineage of a clinically important opportunistic bacterial pathogen, *Pseudomonas aeruginosa*, as it adapts to the airways of several individual cystic fibrosis patients over 200,000 bacterial generations, and provide estimates of mutation rates of bacteria in a natural environment. In contrast to predictions based on in vitro evolution experiments, we document limited diversification of the evolving lineage despite a highly structured and complex host environment. Notably, the lineage went through an initial period of rapid adaptation caused by a small number of mutations with pleiotropic effects, followed by a period of genetic drift with limited phenotypic change and a genomic signature of negative selection, suggesting that the evolving lineage has reached a major adaptive peak in the fitness landscape. This contrasts with previous findings of continued positive selection from long-term in vitro evolution experiments. The evolved phenotype of the infecting bacteria further suggests that the opportunistic pathogen has transitioned to become a primary pathogen for cystic fibrosis patients.**

microbial evolution | natural population | chronic infection | genome sequences | transcriptomics

Bacterial populations maintained in defined in vitro systems are well-suited for experimental evolutionary studies of adaptation to novel environments. It has been observed from such investigations that the initial rate of adaptation is rapid with a continuous increase in fitness that later tends to decelerate over time (1). Experimental evolution has also been used to gain insight into the molecular details of evolution (2–7). Interestingly, a complex relationship between the decelerating rate of evolutionary adaptation and a constant rate of genomic changes enriched with nonsynonymous changes has been observed in one such in vitro experiment (2). The level of diversity that evolves in a population is thought to increase as a function of environmental complexity in terms of available ecological niches. Laboratory experiments with artificially structured microbial populations have shown that environmental heterogeneity, such as population complexity and spatial structure, can generate and maintain population diversity (8–10). Comparable systematic investigations of bacterial populations living in complex natural environments are lacking, and the extent to which laboratory evolutionary experiments bear any relevance to evolutionary processes in natural environments is poorly understood. To address this question, we investigated the evolutionary dynamics including the roles of neutral versus selective mechanisms in bacterial populations evolving over 200,000 generations in the complex natural and structured environment of cystic fibrosis (CF) airways.

Chronic airway infections in CF patients by the bacterium *Pseudomonas aeruginosa* offer a rare opportunity to study evolutionary adaptation of bacteria in natural environments. CF airways, like other natural environments, constitute a complex and highly dynamic ecosystem composed of a range of niches that vary in space and time. Factors such as heterogeneous distribution of oxygen, nutrients, and other substances, fluctuations in exposure to the host immune system and therapeutic antibiotic treatments, as well as a diverse microbial community structure in the host all contribute to environmental conditions under which the bacteria face continuous adaptive challenges (11). These environmental conditions have been used to explain the extensive phenotypic diversity often observed in *P. aeruginosa* populations from the same CF respiratory specimen (12–14). We propose that *P. aeruginosa* infections in CF airways constitute a valuable natural system for addressing central questions concerning evolutionary biology, and we use this system to investigate how the highly successful and transmissible DK2 lineage of *P. aeruginosa* adapts to this environmental complexity in several CF patients using genome sequencing, transcriptional profiling, and phenotypic arrays. Based on previous estimates of in situ growth rates of *P. aeruginosa* isolates (15), we calculate that the DK2 lineage has persisted for about 200,000 bacterial generations in the CF airways of several patients, and as such this study documents the genomic and functional details of a long microbial evolution "experiment."

## Results and Discussion

This study was based on a collection of *P. aeruginosa* isolates sampled from several hundred Danish CF patients between 1973 and 2008. Initial examination of the clonal relationship among a subset of the stored isolates using molecular typing methods revealed that some *P. aeruginosa* clonal lineages were observed repeatedly because of transmission among the patients (16). For the present study, we are focusing on the DK2 lineage (*Materials and Methods*), which we have established as one of these transmissible lineages and isolated from ~40 patients since the start of the sampling program in 1973.

**Genomic Evolution of the DK2 Lineage.** To enable a detailed characterization of the evolutionary trajectory associated with the dissemination of the DK2 lineage across multiple patients, we sequenced the genomes of 12 selected DK2 strains (CF114-1973,

Author contributions: L.J. and S.M. conceived study; L.Y., L.J., M.O.A.S., and S.M. designed research; L.Y., L.J., R.L.M., S.D., M.H.R., and M.O.A.S. performed research; H.K.J., O.C., and N.H. contributed new reagents/analytic tools; L.Y., L.J., R.L.M., S.D., C.T.W., M.H.R., S.K.H., A.F., M.O.A.S., and S.M. analyzed data; and L.Y., L.J., R.L.M., S.D., A.F., M.O.A.S., and S.M. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]L.Y. and L.J. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: sm@bio.dtu.dk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1018249108/-/DCSupplemental.

CF43-1973, CF66-1973, CF66-1992, CF66-2008, CF30-1979, CF173-1984, CF173-2005, CF333-1991, CF333-1997, CF333-2003, and CF333-2007) isolated from 6 different patients between 1973 and 2008 (Fig. 1 and Tables S1 and S2). We identified high-quality SNPs in the nonrepetitive parts of the genomes by mapping sequence reads for each isolate against CF333-2007 (Table S3 and *SI Materials and Methods*). A total of 368 SNPs have appeared since the clones diverged from their common ancestor before 1973 (Table S3), and we find that there are 180 SNPs between the most distantly related clones (CF114-1973 and CF333-2007). Based on the identified SNPs, we construct a tree representing the evolutionary relationship of the DK2 strains (Fig. 2*A* and *Materials and Methods*), and as such the tree demonstrates the sequence of mutational events in the DK2 lineage from 1973 and onward.

We predict the DK2 ancestor to be related to the sequenced strains, as indicated with the predicted root in Fig. 2*A*, based on the assumption that the DK2 ancestor would contain wild-type SNP alleles [i.e., SNP positions with sequences identical to homolog sequences found in the *P. aeruginosa* reference strains PAO1 and PA14 (17)]. Despite the genomic diversity among the 1973 isolates, their phenotypes are similar to the well-characterized laboratory *P. aeruginosa* strain PAO1 (Fig. 3 and Table S1). Because the CF infection process is associated with the appearance of phenotypes of which many are not usually observed among environmental, wild-type isolates of this species (18–20), we predict that the immediate ancestor to the DK2 lineage entered the CF environment shortly before the first sampling in 1973. Because the original infecting DK2 strain is unavailable for analysis, we cannot precisely determine the pre-1973 events that resulted in genomic diversity among the isolates sampled in 1973. Although a complete understanding of the pre-1973 history of the DK2 lineage is not a prerequisite for the present study, we hypothesize that this diversity may be a result of processes such as (*i*) an already-present diversity in the ancestral DK2 lineage before first entry into the CF environment (i.e., the first infection caused by a mix of genetically distinct subpopulations of DK2) and/or a rapid diversification of the ancestral DK2 strain upon first infection as observed in patients infected with environmental strains (13, 19) possibly coupled to a transmission process that retained diversity, (*ii*) the storage conditions of these particular isolates in agar slants at room temperature for an extended period in which genomic changes may have arisen (21), or (*iii*) an increased mutation rate in the lineage before the 1973 sampling. The presence of a missense mutation in *mutY* and a molecular signature consistent with a MutY defect [i.e., a high proportion of transversions ($P = 7 \times 10^{-13}$) of which 85% are G·C→T·A] in the sublineage branching into CF43-1973 indeed demonstrates a pre-1973 hypermutator phenotype for this particular isolate (22). As the two remaining 1973 strains did not show molecular signs of a hypermutator phenotype, other phenomena may have led to increased genomic diversification in these strains before sampling in 1973.

We derive two conclusions based on the topology of the tree. First, all isolates sampled from 1979 and onward share a common branch point (marked with a red circle in Fig. 2*A*). We hypothesize that important mutations accumulated in the early hosts (pre-1979) positioned the DK2 lineage for subsequent reproductive success in the CF airways in several hosts. Second, the tree demonstrates a remarkably limited diversification of the sequential isolates present in patients CF333 and CF66 after 1979. This observation is surprising, because CF airways represent a dynamic and spatially structured environment in which one would expect a divergent population structure and multiple coexisting sublineages resulting in a deeply branched phylogenetic tree of clones isolated from the same host (8, 12, 13, 23). Note, however, the disconnected branching of the 1984 and 2005 isolates in patient CF173 (Fig. 2*A*), which suggests that a replacement of the DK2 lineage caused by a secondary transmission event after 1984 has taken place.

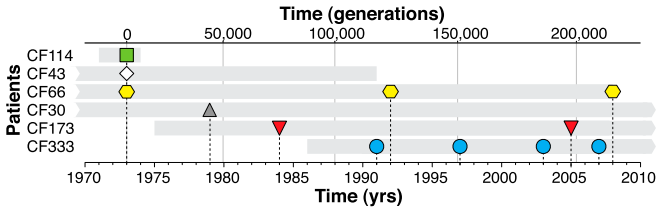


**Fig. 1.** Isolate sampling points and patient life span. DK2 *P. aeruginosa* isolates were collected from six different CF patients during a 35-y time period. Bacterial isolates are represented by the colored symbols, and gray bars represent patient life span.

Because the DK2 lineage has exhibited a highly successful phenotype over an estimated period of 200,000 generations of growth, it provides an excellent system for the detailed study of molecular evolution in ways that previously have been possible only in well-defined artificial laboratory environments. Importantly, we find that SNPs in the protein-coding regions of the genome have accumulated at a constant rate in the DK2 lineages since 1973 (Fig. 2 *B* and *C*). We can estimate the mutation rate based on the accumulation of synonymous mutations in the protein-coding regions of the genome assuming that they are neutral and not selected for (2, 24). Based on the codon use in *P. aeruginosa* (25), we find that 25% of all SNPs in the protein-coding parts of the genome result in synonymous changes (*SI Materials and Methods*). We fit a least-squares linear model of mutation accumulation over time and calculate the mutation rate as the number of synonymous mutations fixed per generation (cf. Fig. 2*B*) divided by the fraction of mutations expected to be synonymous in the genome. Based on these assumptions, we find that the mutation rate of *P. aeruginosa* in CF airways is $7.2 \times 10^{-11}$ SNPs/bp per generation, which is in the range of previous estimates for related bacteria (2, 26, 27). Excluding the 1973 isolates from the calculation, we find the mutation rate to be $6.9 \times 10^{-11}$ SNPs/bp per generation. It is important to note that these estimates show no indication of increased rates of mutation, as has been suggested for biofilm-associated cells and CF infection-associated cells due to oxidative stress or other phenomena (28–32).

**Selective Forces Acting on the DK2 Genome.** Based on the number of protein-coding synonymous SNPs, we can estimate the expected number of nonsynonymous protein-coding SNPs assuming a neutral theory of molecular evolution. This theory predicts the rate of nonsynonymous substitutions (dN) relative to the rate of synonymous substitutions (dS) to approach unity during neutral evolution. Over- or underrepresentation of nonsynonymous mutations can be seen as a sign of positive or negative selection, respectively. Overall, among all mutations in the protein-coding regions of the genome, we find a significant underrepresentation of nonsynonymous mutations (dN/dS = 0.79; $P = 1.3 \times 10^{-4}$) (Table S4). The sign of negative selection is most prevalent in isolates from 1979 and onward (branches G-U in Fig. 2*A*; dN/dS = 0.70; $P = 6.5 \times 10^{-5}$), whereas calculations based on early isolates (branches A–F in Fig. 2*A*) do not support evidence of negative selection (dN/dS = 0.89; $P = 0.10$). This observation supports a hypothesis that a selection of adaptive mutations takes place in the years up to 1979 leading to a highly adapted clone closely related to CF30-1979 (represented by a red circle in Fig. 2*A*). Although our results contrast with previous findings of strong positive selection in CF airways (23), it is likely that the earliest stages (before 1973) of infection by the DK2 lineage were also dominated by positive selection mechanisms. Indeed, despite of the overall sign of negative selection, there is a statistically significant sign of positive selection in the lineage (branch F in Fig. 2*A*) leading to the post-1979 isolates (dN/dS = 1.58; $P = 0.02$).

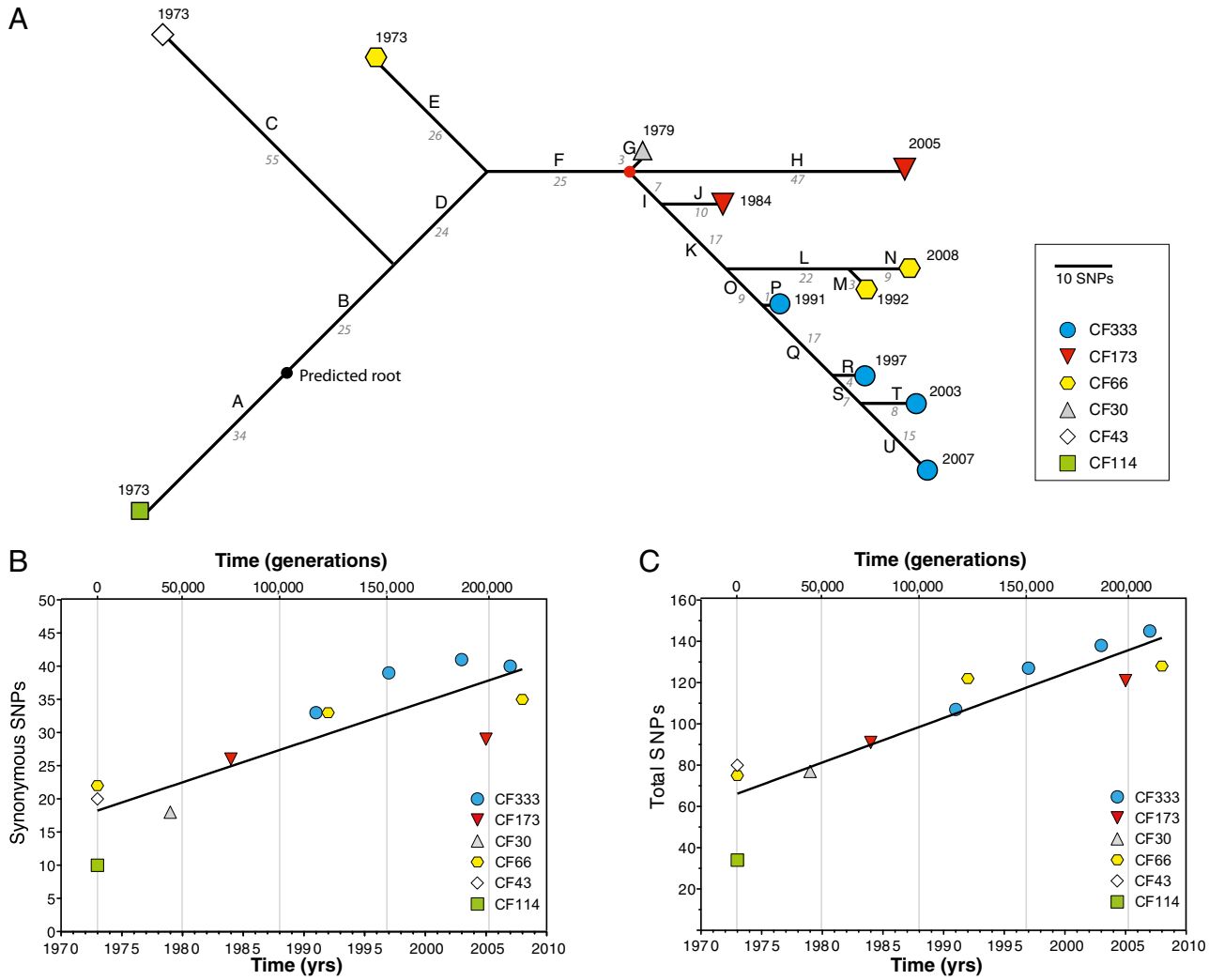


**Fig. 2.** Evolutionary trajectory of the DK2 lineage. (*A*) Tree representing the evolutionary relationship of the DK2 clones. The tree is based on the accumulation of SNPs identified from genome sequencing. Lengths of branches are proportional to the SNPs between the isolates as indicated by the numbers in italics. The black circle that separates branches "A" and "B" represents the predicted root (i.e., the immediate common ancestor of the isolates). Synonymous (*B*) and total (*C*) SNPs accumulated between the predicted root and the tip of each branch are plotted against time (and estimated generations).

**Phenotypic Changes During Long-Term Adaptation to CF Airways.** The dynamics of phenotype changes in the isolates of the DK2 lineage was characterized both by gene expression profiles under a single growth condition (standard laboratory medium) and by determination of catabolic activities under different growth conditions. Each method interrogates different aspects of cellular physiology (whole-genome transcript abundance and cellular catabolic capacity) and is therefore expected to probe different components of the genotype/phenotype relationship of the bacteria.

Hierarchical clustering of all recorded expression and catabolic signals from each isolate showed that strains sampled between 1979 and 2008 are more similar to each other than to the initial isolates (1973) (Fig. S1 *A* and *B*). To further examine the extent of gene expression changes among the isolates a principal component analysis, which transforms a complex data matrix into a small number of variables, was applied. The first principal component PC1 accounts for most (42%) of the variance in the dataset. Plotting of the loadings of each isolate on PC1 as a function of time shows a strikingly simple and clear time-dependent progression, with a major shift between 1973 and 1979 followed by a period with few changes (Fig. 3*A*). After 1979 all of the DK2 isolates share a very similar gene expression pattern despite the fact that they cover samples from four different patients taken with a time span of 29 y (estimated to correspond to ~150,000 bacterial generations) and have accumulated 179 mutations.

In analogy with the global transcriptional data, the "catabolome" datasets display a first principal component diagram that documents a lack of variation among strains isolated after 1979 (Fig. 3*B*). Before 1979 the profile changed rapidly from a wild-type pattern of catabolism of C and N sources, as observed in a reference strain (PAO1) and in the 1973 isolates, to the adapted pattern reflecting a mutation(s) in a global regulatory gene(s). Surprisingly, the catabolome as monitored by the Biolog phenotype arrays shows only loss of function (Fig. S2), which is paradoxical, taking the abundance of, for example, amino acids in the sputum into consideration (33). We hypothesize that the airway mucus and the surrounding environment is rich in a variety of nutrients and that the competition for these from other infecting species is marginal, providing ample opportunities for alternative feeding by *P. aeruginosa*. Overall, our phenotype data are consistent with our genomic data and suggest strong negative selective forces acting on the evolving populations of the DK2 lineage after 1979.

The observed phenotypic stability and the genomic signature of negative selection are consistent with a hypothesis that the

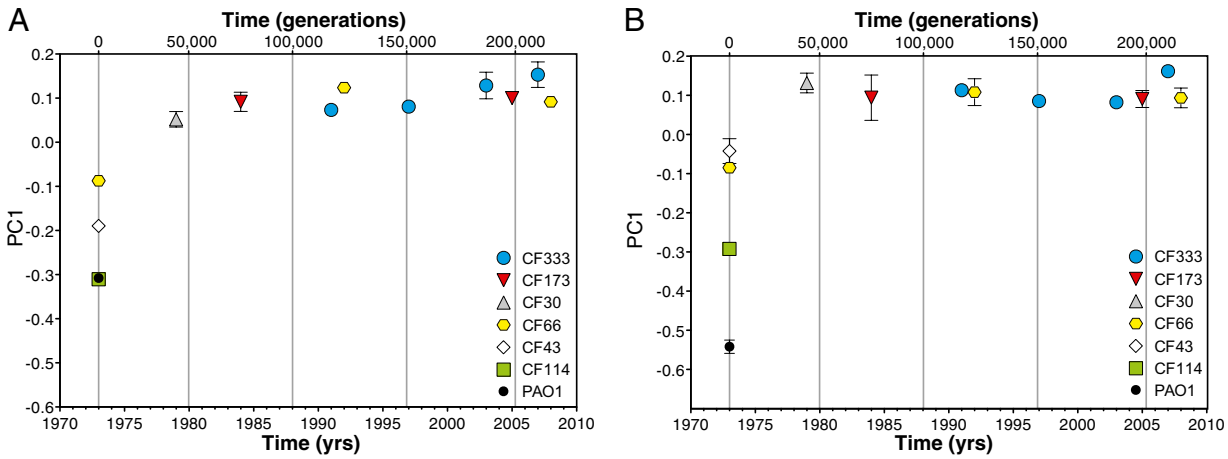


**Fig. 3.** Phenotypic relationships and dynamics among DK2 isolates. Principal component analyses of the microarray expression data (*A*) and Biolog phenotypic data (*B*) show a similar pattern suggesting that major changes occurred before 1979. First principal component (PC1) loadings for each isolate are graphed as a function of time (and estimated generations). PC1 represents 42% (*A*) and 63% (*B*) of the total variation among isolates. Errors bars indicate standard deviations of biological triplicates (*A*) or duplicates (*B*) of the same isolate.

DK2 lineage (represented by the evolved genotype/phenotype of CF30-1979 and later isolates) becomes positioned on a major fitness peak in this complex and dynamic natural environment. If the evolved DK2 lineage is in fact positioned on a major fitness peak, most new mutations would result in reduced fitness and therefore not be fixed in the population (negative selection), whereas neutral (synonymous) mutations would be fixed (genetic drift), which is in agreement with our findings.

**Parallel Evolution in DK2 Sublineages.** The striking similarity between the phenotypes of DK2 isolates sampled from different CF patients suggests that the host environments represent parallel selective conditions by which evolution is directed. Such parallelism is expected to generate parallel evolution (20), and a search for signs of independent evolution of the same traits in the isolates confirms this. In the case of two phenotypes that evolved independently in the CF173 and CF333 isolates after their early separation from CF30-1979, we were able to identify the genetic details: (*i*) The lost capacity of both isolates to catabolize 4-hydroxyphenylacetic acid (4-HPA), a trait identified in the Biolog phenotypic arrays, is associated with similar—but not identical—genetic changes (Fig. 4 *A*–*C*); and (*ii*) increased resistance to the antibiotic ciprofloxacin conferred by two independent *gyrA* mutations that have previously been described to contribute to ciprofloxacin resistance (34) (Fig. 4*D*). Consistent with this, both isolates CF173-2005 and CF333-2007 show increased levels of resistance toward ciprofloxacin (Fig. 4*D*). Additionally, both isolates obtained mutations in the genes PA5160 and *mexY*, which are components of multidrug transport systems (35) (Fig. 4*A*).

**Mutations with Pleiotropic Effects Drive Adaptation to New Environments and Reproductive Success in CF Airways.** The genomic analysis showed that mutations appeared in several global regulatory genes during the initial 42,000 estimated generations of growth. Indeed, we observed more missense and nonsense SNPs in regulatory genes in the first 42,000 estimated generations before 1979 (11 mutations) compared with the following 167,000 estimated generations after 1979 (5 mutations) (Table S5). Of particular interest are mutations in the *mucA* (*SI Materials and Methods*) and *rpoN* genes that occurred before 1973 (as observed in CF66-1973), and a deletion of the *lasR* gene that occurred before 1979 (as observed in CF30-1979). Gene expression controlled by MucA, LasR, and RpoN accounts for approximately half of the total significantly changed gene expression between CF114-1973 and CF30-1979, which confirms the pleiotropy of these mutations in the DK2 lineage (*SI Materials and Methods*). Notably, transcomplementation of late DK2 isolates with wild-type RpoN expressed from a recombinant plasmid restored to a large extent the catabolic profile of wild-type *P. aeruginosa* (Fig. S2), suggesting that much of the adapted catabolome activity in DK2 is directly associated with the *rpoN* mutation already present in the CF66-1973 isolate. After 1979 no similarly dramatic global regulatory changes seem to have occurred in the DK2 lineage.

Based on the phenotypic data described above and tracking of the mutated alleles of *mucA*, *lasR*, and *rpoN* among DK2 isolates, we conclude that all DK2 clones isolated after 1979 from more than 40 different patients have a common ancestor with a genotype and phenotype like the one described for the CF30-1979 isolate (*SI Materials and Methods*). These findings show that extensive transmission of the DK2 lineage has occurred in the

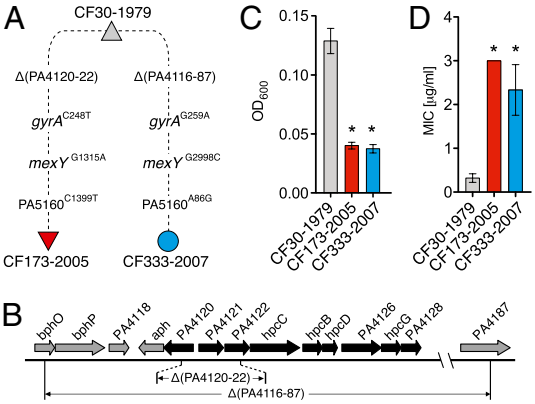


**Fig. 4.** Parallel evolution in CF333 and CF173 isolates. (*A*) Schematic representation of the four parallel, independent genetic changes observed in the DK2 lineage in CF333 and CF173. The genetic changes include missense/nonsense mutations in three genes (*gyrA*, *mexY*, and PA5160) and deletions of genes involved in 4-HPA catabolism. (*B*) Outline of the chromosomal region with genes required for 4-HPA catabolism (labeled in black) and positions of the deletions Δ(PA4116-87) (CF333) and Δ(PA4120-22) (CF173). (*C*) Growth of strains in minimal medium with 4-HPA as carbon source. (*D*) Ciprofloxacin minimal inhibitory concentration (MIC; μg/mL) values measured using Etest strips (AB Biodisk). Error bars represent standard deviations of three replicate experiments. Asterisks show significant differences ($P < 0.05$).

clinic after the occurrence of the last (*lasR*) of the three global regulatory mutations. Interestingly, in CF66, in whom the original DK2 strain found in 1973 went extinct (either by successful treatment or by competition with other genotypes of *P. aeruginosa* or both), a later transmission of an adapted DK2 lineage (the CF30-1979 adapted phenotype) occurred in 1992. This further supports that reproductive success (fitness) of the DK2 lineage was dependent on the particular "genetic configuration" found in CF30-1979, and that components of this fitness can be described by the phenotypes measured here.

The infecting bacteria in CF airways represent a continuous challenge to the immune system, and it is striking that among the many transcriptional changes imposed by the three global regulatory mutations (*mucA*, *rpoN*, *lasR*), reduced gene expression in the DK2 isolates is in fact associated with *P. aeruginosa* antigenic virulence factors (19, 36, 37). It is well-known that *lasR* and *rpoN* regulate virulence in this organism, and recently we documented that also *mucA* exerts control of expression of several virulence genes (19). It is therefore tempting to speculate that the combined reduction of virulence caused by the three early regulatory mutations is a key factor in the success of the DK2 lineage of *P. aeruginosa*. It should also be noted that very few regulatory mutations are identified in the lineage after 1979 (Table S5).

## Concluding Remarks

We have presented a detailed analysis of the temporal sequence of all of the genetic and phenotypic changes that lead to the successful establishment of a bacterial lineage in a large number of new parallel environments. The resulting map of evolutionary changes shows that the DK2 lineage underwent an initial period of rapid adaptation upon colonization of the cystic fibrosis airways followed by a long period with limited phenotypic change and a genomic signature of negative selection.

This study of bacterial evolution in a natural environment challenges the generality of two major findings from laboratory evolution experiments. First, our data show that successful and adapted bacterial populations in natural systems can have genomic signatures of negative selection, in contrast with previous findings of continuous positive selection from evolving populations in constant, nonstructured in vitro environments (2). We note, however, that the earliest stages after the DK2 lineage migrated from the environment to the CF airways (before 1973) may have been dominated by positive selection. Second, our data demonstrate that a complex and variable environment can be dominated by a highly homogeneous population, in contrast to previous findings of diversity promotion and maintenance from laboratory experiments (8–10). These findings underscore the importance of studying the molecular details of evolving populations in natural environments in addition to populations in laboratory systems to further our understanding of basic evolutionary processes.

*P. aeruginosa* is an opportunistic pathogen capable of switching lifestyle as it moves from the environment to compromised hosts. The successful adaptation of the DK2 lineage in a group of CF patients in the Copenhagen CF Clinic is associated with severely reduced growth rate, loss of substantial catabolic activities and motility, and inactivation of important regulatory functions (Table S1). It is unlikely that these changes would not affect the organism's fitness in its original environment. In CF airways, however, the DK2 lineage has proven to be persistent, dominant, and transmissible. The overall adapted phenotype of DK2 is therefore one of a true pathogen, and its evolutionary history illustrates the process of how an opportunistic pathogen with a broad environmental range can transition to a primary-host-specific pathogen for CF patients. With the presented map of evolutionary changes behind this shift, it may now become possible to design novel therapeutic approaches that specifically target the adaptive process and in that way block development of chronic CF lung infection.

## Materials and Methods

**Bacterial Isolation.** *P. aeruginosa* isolates sampled from several hundred Danish CF patients since 1973 have been stored in the Copenhagen Cystic Fibrosis Clinic. Isolation and identification of *P. aeruginosa* from sputum was done as previously described (16). The number of *P. aeruginosa* cfu in such respiratory secretions may be as high as $10^7$–$10^9$ per mL (38). We confirmed that all strains analyzed in the present study belong to the same lineage DK2 [formerly the "b" lineage (16)] using three independent methods for determination of clonal relationships: pulsed-field gel electrophoresis, SNP typing using an ArrayTube microarray (39), and multilocus sequence typing. Single colonies were cultivated and stored at −80 °C with glycerol. The initial storage procedure for the earliest clinical isolates (1973) involved storage in agar slants at room temperature for an extended period before storage at −80 °C.

**Genome Sequencing.** Genomic DNA was prepared from *P. aeruginosa* isolates using a Wizard Genomic DNA Purification Kit (Promega). Genome sequences of CF333-2007 and CF333-1991 were sequenced by 454 pyrosequencing with an approximate coverage of 20-fold. Reads were de novo assembled into contigs by Newbler version 1.1.03.24 (Roche) resulting in 164 and 87 contigs, respectively, that comprised 6,340,400 and 6,350,334 bp, respectively, and the two genome sequences was aligned using MUMmer3 (40) resulting in an alignment covering 6,336,193 bp of the genomes. Isolates CF114-1973, CF43-1973, CF66-1973, CF66-1992, CF66-2008, CF30-1979, CF173-1984, CF173-2005, CF333-1997, CF333-2003, and CF333-2007 were sequenced on an Illumina GAII generating 75-bp single-end reads using a multiplexed protocol to an average coverage of 47-fold (Table S2). The reads from each isolate were mapped against the reference genome of CF333-2007 using Bowtie release 0.12.3 (41), and SNP variants were called using SAMtools release 0.1.7 (42).

**Construction of the Phylogenetic Tree.** Genome sequences of all isolates show exhaustive homology to the reference CF333-2007 with no genome alignments covering less than 97.6% of the reference genome sequence (Table S2). Based on the SNP mutations identified in the alignments, we constructed a phylogenetic tree representing the evolutionary relationship among the DK2 strains. The branches at which the individual SNP mutations accumulated were determined by comparing the SNP mutations among the isolates. For example, if a mutation was found in isolates CF333-1997, CF333-2003, and CF333-2007, then the particular SNP mutation was assigned to the branch labeled "Q" in Fig. 2*A*. Except for two cases, all SNP mutations accumulated in a parsimonious fashion, such that a consistent tree could be constructed that shows the sequence of mutational events. This means that if an SNP mutation is shown to have accumulated on a particular branch, for example, branch K, then all clones descending from this branch (CF66-1992, CF66-2008, CF333-1991, CF333-1997, CF333-2003, and CF333-2007) contain the mutated allele. The only two SNP mutations that exhibit an aberrant evolutionary distribution that does not fit the observed sequential accumulation were found in *ampC* and *ftsI*. In the case of *ftsI*, a G188A substitution was found in isolates CF173-2005, CF66-1992, CF66-2008, CF333-1991, CF333-1997, CF333-2003, and CF333-2007 but not in CF173-1984 (Fig. 2*A*). The same ambiguity applies to *ampC*, in which an SNP mutation (A320G) was found in CF173-2005, CF66-1992, CF66-2008, CF333-2003, and CF333-2007 but not in CF333-1991, CF333-1997, and CF173-1984 (Fig. 2*A*). The irregular distribution of these particular SNP mutations may be a result of identical, but independent, mutational events, direct reversions, or horizontal transfer of DNA encoding the SNP. Notably, both *ampC* and *ftsI* have been reported to be under antibiotic-associated selection pressure (43, 44).

As the *ampC* and *ftsI* loci are the only two cases that confer ambiguity, we conclude that the tree structure is robust, although we cannot exclude the possibility that other similar events may be undetectable in the dataset. Because the immediate ancestor to the DK2 lineage is unavailable, a precise determination of the root of the tree is not possible. We predict the DK2 ancestor to be related to the sequenced strains as indicated by the predicted root in Fig. 2*A* based on the assumption that the SNP positions in the DK2 ancestor would be identical in sequence to the homolog sequences found in other *P. aeruginosa* reference strains such as PAO1, PA14, and LESB58.

**Estimation of in Situ Bacterial Generations.** The estimation of in situ generations of the DK2 lineage from 1973 to 2008 is based on our previous determination of in vivo growth rates of *P. aeruginosa* populations within sputum material sampled from different CF patients (15). Considering that growth rates of *P. aeruginosa* decrease when infection progresses in CF patients, we segregated the time period from 1973 to 2008 into several intervals and estimated the in

situ growth rate of each interval according to the in vitro growth rate. Generations of each isolates were calculated as sum of generations from 1973 to the isolation years (see details in Table S1 and *SI Materials and Methods*).

**Global Phenotypic Profiling.** Transcriptomes were measured using Affymetrix GeneChip for *P. aeruginosa*. Phenotype microarrays PM1, PM2, and PM3 (Biolog) were used to obtain the metabolic profiles of different carbon and nitrogen sources. Analysis of both types of datasets was performed using BioConductor software (see *SI Materials and Methods* for details) (45).

**ACKNOWLEDGMENTS.** We acknowledge the expert assistance of Francois Vigneault and George M. Church in genome sequencing. The project was supported by grants from the Danish Council for Independent Research/ Natural Sciences and from The Lundbeck Foundation.

1. Elena SF, Lenski RE (2003) Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. *Nat Rev Genet* 4:457–469.
2. Barrick JE, et al. (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461:1243–1247.
3. Shendure J, et al. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309:1728–1732.
4. Herring CD, et al. (2006) Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat Genet* 38: 1406–1412.
5. Cooper TF, Rozen DE, Lenski RE (2003) Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc Natl Acad Sci USA* 100: 1072–1077.
6. Zambrano MM, Siegele DA, Almirón M, Tormo A, Kolter R (1993) Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. *Science* 259:1757–1760.
7. Giraud A, et al. (2008) Dissecting the genetic components of adaptation of *Escherichia coli* to the mouse gut. *PLoS Genet* 4:e2.
8. Rainey PB, Travisano M (1998) Adaptive radiation in a heterogeneous environment. *Nature* 394:69–72.
9. Boles BR, Thoendel M, Singh PK (2004) Self-generated diversity produces "insurance effects" in biofilm communities. *Proc Natl Acad Sci USA* 101:16630–16635.
10. Hansen SK, Rainey PB, Haagensen JA, Molin S (2007) Evolution of species interactions in a biofilm community. *Nature* 445:533–536.
11. Lyczak JB, Cannon CL, Pier GB (2002) Lung infections associated with cystic fibrosis. *Clin Microbiol Rev* 15:194–222.
12. Wilder CN, Allada G, Schuster M (2009) Instantaneous within-patient diversity of *Pseudomonas aeruginosa* quorum-sensing populations from cystic fibrosis lung infections. *Infect Immun* 77:5631–5639.
13. D'Argenio DA, et al. (2007) Growth phenotypes of *Pseudomonas aeruginosa* lasR mutants adapted to the airways of cystic fibrosis patients. *Mol Microbiol* 64:512–533.
14. Martin DW, et al. (1993) Mechanism of conversion to mucoidy in *Pseudomonas aeruginosa* infecting cystic fibrosis patients. *Proc Natl Acad Sci USA* 90:8377–8381.
15. Yang L, et al. (2008) In situ growth rates and biofilm development of *Pseudomonas aeruginosa* populations in chronic lung infections. *J Bacteriol* 190:2767–2776.
16. Jelsbak L, et al. (2007) Molecular epidemiology and dynamics of *Pseudomonas aeruginosa* populations in lungs of cystic fibrosis patients. *Infect Immun* 75: 2214–2224.
17. Winsor GL, et al. (2009) *Pseudomonas* Genome Database: Facilitating user-friendly, comprehensive comparisons of microbial genomes. *Nucleic Acids Res* 37(Database issue):D483–D488.
18. Nguyen D, Singh PK (2006) Evolving stealth: Genetic adaptation of *Pseudomonas aeruginosa* during cystic fibrosis infections. *Proc Natl Acad Sci USA* 103:8305–8306.
19. 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* 12:1643–1658.
20. Huse HK, et al. (2010) Parallel evolution in *Pseudomonas aeruginosa* over 39,000 generations in vivo. *MBio* 1:e00199–10.
21. Naas T, Blot M, Fitch WM, Arber W (1994) Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. *Genetics* 136:721–730.
22. Nghiem Y, Cabrera M, Cupples CG, Miller JH (1988) The mutY gene: A mutator locus in *Escherichia coli* that generates G.C—T.A transversions. *Proc Natl Acad Sci USA* 85: 2709–2713.
23. Smith EE, et al. (2006) Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc Natl Acad Sci USA* 103:8487–8492.
24. Kimura M (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ Press, Cambridge, UK).
25. Grocock RJ, Sharp PM (2002) Synonymous codon usage in *Pseudomonas aeruginosa* PA01. *Gene* 289:131–139.
26. Drake JW (1991) A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci USA* 88:7160–7164.
27. Ochman H, Elwyn S, Moran NA (1999) Calibrating bacterial evolution. *Proc Natl Acad Sci USA* 96:12638–12643.
28. Driffield K, Miller K, Bostock JM, O'Neill AJ, Chopra I (2008) Increased mutability of *Pseudomonas aeruginosa* in biofilms. *J Antimicrob Chemother* 61:1053–1056.
29. Ciofu O, Riis B, Pressler T, Poulsen HE, Høiby N (2005) Occurrence of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis patients is associated with the oxidative stress caused by chronic lung inflammation. *Antimicrob Agents Chemother* 49: 2276–2282.
30. Boles BR, Singh PK (2008) Endogenous oxidative stress produces diversity and adaptability in biofilm communities. *Proc Natl Acad Sci USA* 105:12503–12508.
31. Oliver A, Canton R, Campo P, Baquero F, Blazquez J (2000) High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* 288: 1251–1254.
32. García-Castillo M, et al. (2010) Stationary biofilm growth normalizes mutation frequencies and mutant prevention concentrations in *Pseudomonas aeruginosa* from cystic fibrosis patients. *Clin Microbiol Infect*, 10.1111/j.1469-0691.2010.03317.x.
33. Barth AL, Pitt TL (1996) The high amino-acid content of sputum from cystic fibrosis patients promotes growth of auxotrophic *Pseudomonas aeruginosa*. *J Med Microbiol* 45:110–119.
34. Kureishi A, Diver JM, Beckthold B, Schollaardt T, Bryan LE (1994) Cloning and nucleotide sequence of *Pseudomonas aeruginosa* DNA gyrase gyrA gene from strain PAO1 and quinolone-resistant clinical isolates. *Antimicrob Agents Chemother* 38: 1944–1952.
35. Stover CK, et al. (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 406:959–964.
36. Dasgupta N, et al. (2003) A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in *Pseudomonas aeruginosa*. *Mol Microbiol* 50:809–824.
37. Hentzer M, et al. (2003) Attenuation of *Pseudomonas aeruginosa* virulence by quorum sensing inhibitors. *EMBO J* 22:3803–3815.
38. Singh PK, et al. (2000) Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature* 407:762–764.
39. Wiehlmann L, et al. (2007) Population structure of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* 104:8101–8106.
40. Kurtz S, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5:R12.
41. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.
42. Li H, et al. (2009) 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
43. Spangenberg C, Montie TC, Tümmler B (1998) Structural and functional implications of sequence diversity of *Pseudomonas aeruginosa* genes oriC, ampC and fliC. *Electrophoresis* 19:545–550.
44. Gotoh N, Nunomura K, Nishino T (1990) Resistance of *Pseudomonas aeruginosa* to cefsulodin: Modification of penicillin-binding protein 3 and mapping of its chromosomal gene. *J Antimicrob Chemother* 25:513–523.
45. Gentleman RC, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.