Title

Genetic adaptation and transmission of *Achromobacter* in cystic fibrosis

Authors and affiliations

Migle Gabrielaite[1,*], Niels Nørskov-Lauritsen[2], Finn C. Nielsen[1], Helle K. Johansen[3,4], Rasmus L. Marvig[1]

[1] Center for Genomic Medicine, Rigshospitalet, Copenhagen, Denmark

[2] Department of Clinical Microbiology, Aarhus University Hospital, Aarhus Denmark

[3] Department of Clinical Microbiology, Rigshospitalet, Copenhagen, Denmark

[4] Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[*] Correspondence to Migle Gabrielaite, migle.gabrielaite@regionh.dk

Abstract

Achromobacter species are increasingly being detected in patients with cystic fibrosis (CF), and this emerging pathogen has been associated with antibiotic resistance and more severe disease outcomes. Nonetheless, little is known about the extent of transmission and genetic adaptation in *Achromobacter* infections.

We sequenced the genomes of 101 clinical isolates of *Achromobacter* collected from 51 patients with CF—the largest longitudinal dataset to-date. We performed a comprehensive pathogenomic analysis to identify pathogen population structure, within-host adaptation, mutational signatures, patient-to-patient transmission events, and associated genetic variation with antibiotic resistance phenotypes.

We found that the population of *Achromobacter* isolates was composed of five different species where *A. xylosoxidans* accounted for 52% of the infections. Most patients were infected by unique *Achromobacter* clone types; nonetheless, patient-to-patient transmission events identified by shared clone types were observed in 35% (N=18) of patients. We found that the same regulatory and inorganic ion transport genes were frequently mutated in persisting clone types within and between species indicating convergent genetic adaptation. Genome-wide association study (GWAS) of six antibiotic resistance phenotypes revealed that the majority of associated genes were involved in transcription and inorganic ion transport.

Overall, we provide insight into pathogenomics of chronic *Achromobacter* infections and show the relevance of whole genome sequencing of clinical isolates. Our findings on evolution and genetic adaptation can facilitate the understanding of disease progression, inform antibiotic treatment, and identify patient-to-patient transmission.

Keywords: within-host evolution; host-pathogen interaction; microbial genomics; genomic adaptation; *Achromobacter*; cystic fibrosis airway infection; pathogen transmission; GWAS

Introduction

The majority of patients with CF are affected by bacterial airway infections which persist for years and often are the cause of respiratory failure and premature death. (1) Analysis of pathogen genomes, i.e. pathogenomics, have shown that within-host pathogen genetic adaptation and transmission plays a role in these infections (2,3), and pathogenomics of the established CF pathogens are widely studied, in particular, in *Pseudomonas aeruginosa*, *Burkholderia cenocepacia*, *B. multivorans* and *Staphylococcus aureus*. (4,5) Nonetheless, there is limited knowledge about other CF airway infection-causing bacteria, such as *Achromobacter* (4,6) which is a rapidly emerging opportunistic pathogen causing chronic infections in patients with CF. (7,8) Understanding of within-host adaptation and transmission is crucial for urgently needed new treatment strategy development, effective pathogen management and elimination. (4,9) For example, *A. ruhlandii* Danish Epidemic Strain (DES) has already been defined as a hypermutable and exceptionally antibiotic-resistant clone type that has spread among Danish patients (10–12).

To investigate the pathogenomics of *Achromobacter*, we sequenced and analyzed the genomes of 101 clinical isolates of *Achromobacter* from 51 patients with CF. We compared the genomes to identify the genetic relationship of isolates and to make a catalog of genetic variation, including both small variant and gene presence/absence analysis, that differentiated the isolates. (13–15) Additionally, we compared our findings with studies on *P. aeruginosa* infections in CF to provide a genomic basis to understanding the differences in infections across pathogen species, which is e.g. relevant to inform if same treatment strategies should be applied. Moreover, by employing GWAS we evaluated how antibiotic resistance profiles can be associated with genetic markers. In summary, this analysis adds to our knowledge about the main *Achromobacter* genomic changes introduced during chronic infections in

patients with CF, ultimately leading to the possibility of genomic-based disease progression prediction and improved strategies to track and treat persistent airway infections.

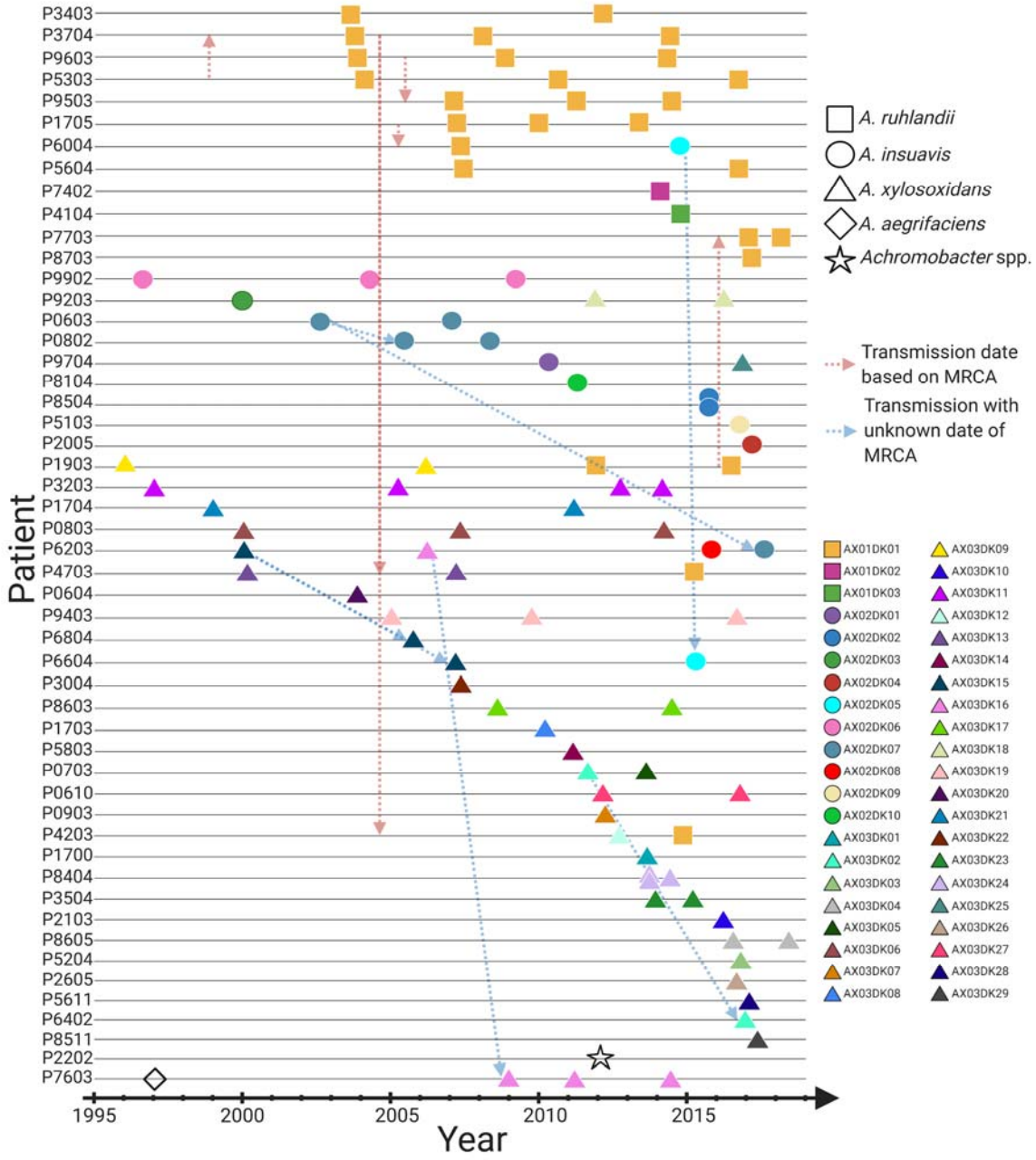

Figure 1. Overview of 101 longitudinally collected *Achromobacter* isolates from patients with CF.

Results

Population structure—five different species

The genomes of 101 *Achromobacter* isolates from the airways of 51 patients with CF were sequenced to follow the within-host evolution and genetic adaptation of the lineages over the initial 0–18 years of infection (Figure 1). All isolates prior to this study were identified in the routine clinical microbiology laboratory by MALDI-TOF or API N20 typing to belong to *A. xylosoxidans* species.

We first compared the sequenced genomes with eight publicly available complete *Achromobacter* reference genomes. While the reference genomes were annotated as *A. xylosoxidans* in the RefSeq database, average nucleotide identity (ANI) analysis with the identity threshold of 95% showed that the eight reference genomes represented four different species (Table S1). We then conducted further ANI analysis and phylogenetic analysis on all *Achromobacter* genomes available in the RefSeq database (141 samples, Table S2) together with clinical *Achromobacter* isolates to identify how our dataset is representative of the overall *Achromobacter* genomes. Our phylogenetic analysis (Figure 2A) revealed that *Achromobacter* annotation is inconsistent among the RefSeq genomes and requires correction to improve species designation (suggested corrections in Table S2).

The genomes of our clinical isolates clustered into five species when using ANI threshold of 95%: *A. ruhlandii*, *A. xylosoxidans*, *A. insuavis*, *A. aegrifaciens* and a new genogroup (Figure 2B). In further analysis, each species was analyzed separately using genetically closest complete reference genome (see Materials and Methods for detailed information).

Out of 51 patients, 15 (25%) were infected with *A. ruhlandii*, 12—with *A. insuavis* (20%), 31—with *A. xylosoxidans* (52%), and 2—with other *Achromobacter* species. Furthermore, isolates from 13 out of 15 patients infected with *A. ruhlandii* species belonged to the previously defined

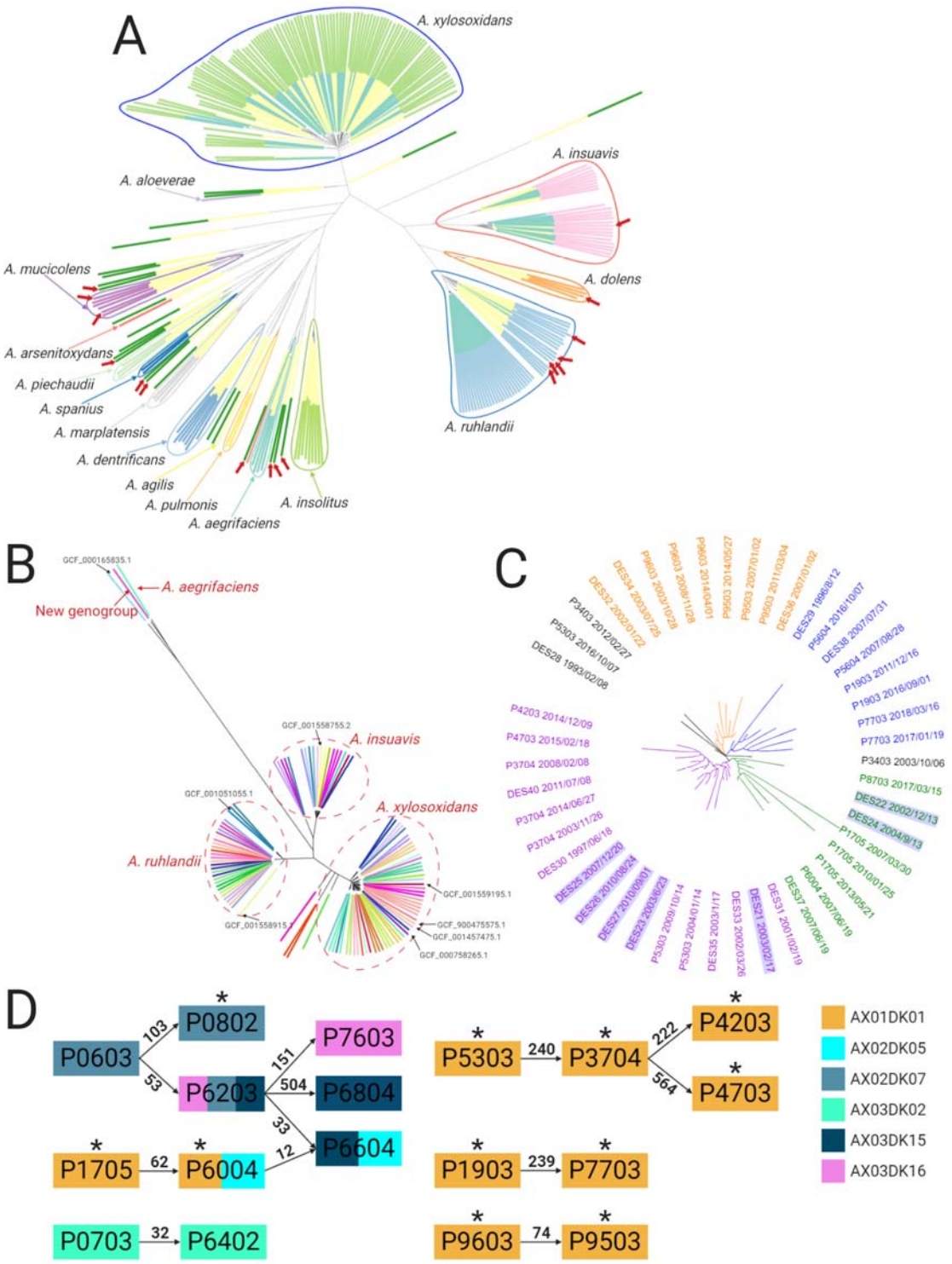

Figure 2. Phylogenies of *Achromobacter* genomes. (A) Phylogenetic tree based on core genome SNPs of 101 *Achromobacter* isolates from patients with CF (inner layer of colors: turquoise) and 141 *Achromobacter* isolates from RefSeq database (inner layer of colors: yellow). Outer layer of colors corresponds to species annotation with suggested corrections; red arrows mark supposedly incorrect species annotation in RefSeq isolates. The

phylogenetic tree can be accessed on Microreact webserver (70). (B) Phylogenetic tree of 101 *Achromobacter* clinical isolates. Colors represent bacterial isolates from different patients; black arrows point to *Achromobacter* reference genomes which were included in the phylogenetic analysis. The phylogenetic tree can be accessed on Microreact webserver (71). (C) Core genome SNP-based phylogenetic tree from *de novo* assembled DES isolates of which 7 were from patients attending CF center in Aarhus (blue label background; (11)) and 25 were from patients attending CF center in Copenhagen. The four patient clusters are separated by branch and label colors. The phylogenetic tree can be accessed on iTOL webserver (72). (D) All identified putative patient-to-patient transmission events, the smallest pairwise distances between the isolates are stated above the arrows. Hypermutator lineages are marked with an asterix. AX01—*A. ruhlandii*, AX02—*A. insuavis* and AX03—*A. xylosoxidans* isolates.

DES clone type (Figure 2C). Finally, 10 out of 51 patients were infected with more than one clone type (N=2) or species (N=9) of *Achromobacter* (Figure 1). The *A. aegrifaciens* and the new genogroup isolates were both genetically distant from all other isolates, and were excluded from further analysis. The remaining 99 *Achromobacter* isolates were grouped into 61 lineages: i.e., all isolates of the same species and the same clone type isolated from the same patient (1–4 isolates per lineage).

Aggregated pan-genome analysis

The aggregated pan-genome was constructed from pan-genomes of each of the 61 lineages (35 of which were single-isolate lineages). This approach allowed us to account for the nature of the dataset where multiple clonal isolates from the same patient are available. The aggregated pan-genome consisted of 21,198 genes: 2,887 core genes, 18,311 accessory genes of which 6,917 genes were unique to a single lineage.

The aggregated pan-genome (Figure 3A) contained *Achromobacter* species-specific genes (649 for *A. ruhlandii*, 648 for *A. insuavis*, and 494 for *A. xylosoxidans*) present in all isolates of the respective species but not in the isolates from other species which further supported the three species separation. Pan-genomes for each *Achromobacter* species were defined by using

all available bacterial isolates available (18–52) for the species. The size of the species' pan-genomes contained 7,070–14,833 genes, of which 4,225–5,130 were core genes, 1,940–10,608 accessory genes and 976–3,162 isolate-unique genes (Table 1).

Table 1. Pan-genome size, number of core, accessory and unique genes for isolates from each Achromobacter species.

Species	Number of bacterial isolates	Pan-genome size	Core genes	Accessory genes	Unique genes
A. ruhlandii	29	7,070	5,130	1,940	976
A. insuavis	18	9,900	4,799	5,101	1,124
A. xylosoxidans	52	14,833	4,225	10,608	3,162

Achromobacter transmission between patients

For all three *Achromobacter* species, our genomic analysis identified events of patient-to-patient transmission. Transmission events were identified based on pairwise SNP distance and phylogenetic relationships of isolates. We furthermore attempted to estimate the date of the most recent common ancestor (MRCA) for all suspected transmission events (Figure 1; Table S3). In total, 13 transmission events were identified with one, two, and three transmitted clone types in *A. ruhlandii*, *A. insuavis* and *A. xylosoxidans*, respectively, among 16 patients (Figure 1). It was previously reported that DES clone type (AX01DK01) has spread among Danish patients with CF (11), and we were able to identify several more recent transmission events between patients. Isolates where patient-to-patient transmission was suspected had relatively small SNP distances (Figure 2D; Table S4) and/or a phylogenetic relationship (Figure S1) in which isolates from one patient are within the clade of the isolates from another patient. Ultimately, we observed that patients were always infected with unique clone types if no transmission between patients was suspected (DES isolates were assumed to have spread through patient-to-patient transmission).

Overall, we observed high *Achromobacter* isolate diversity between and within patients in 26 longitudinally collected lineages: 0–1,034 SNP differences (average: 74 SNPs, median: 15 SNPs) were identified within lineages where late isolates contained 3-fold more SNPs than early isolates (Figure 3B; Table S5).

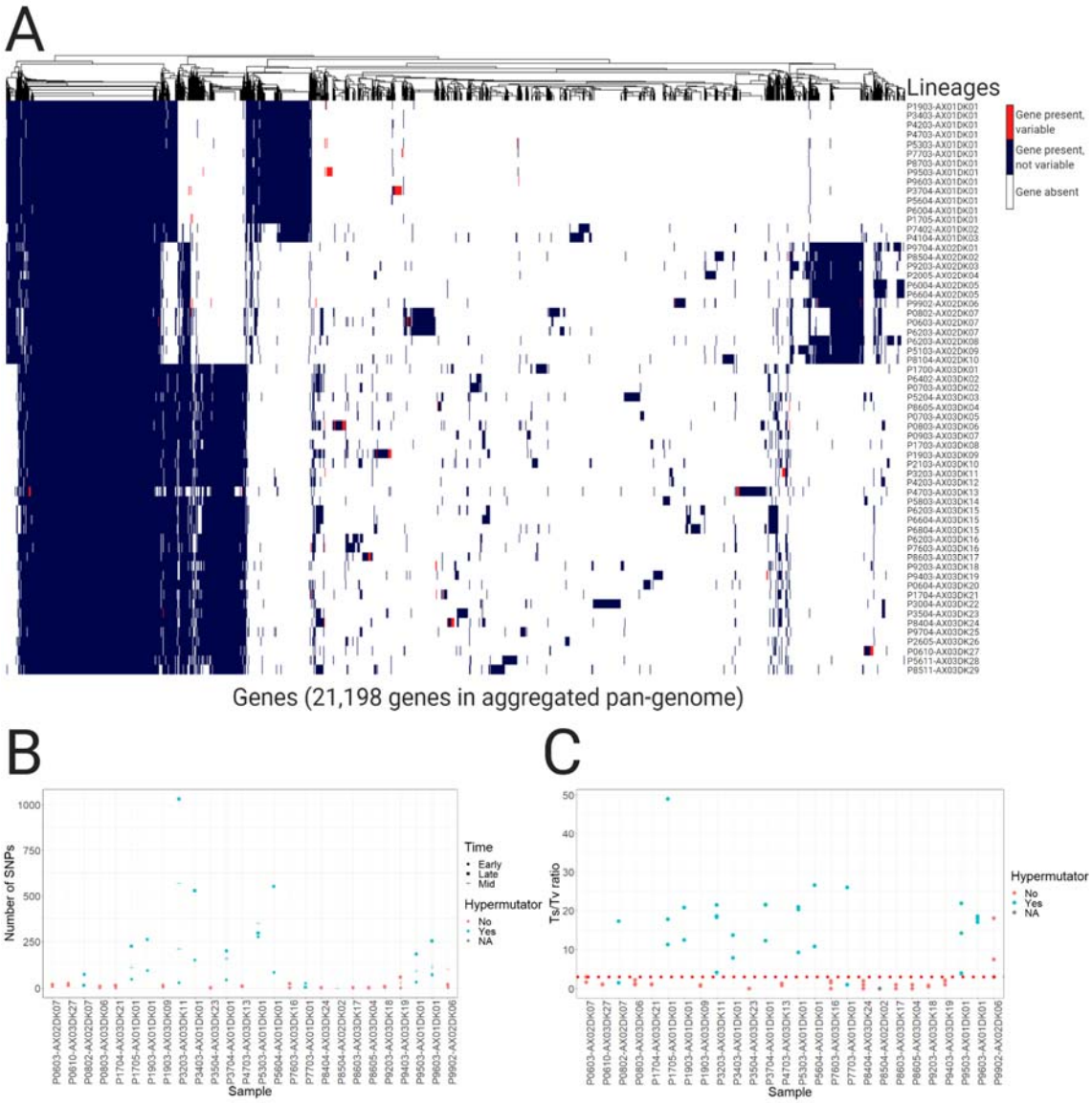

Figure 3. Within- and between-lineage genome variability. (A) The aggregated pan-genome of 61 *Achromobacter* lineages containing gene presence, absence, and gene variability (if the gene is present in some isolates while absent in other isolates within the lineage) information, (B) SNP distribution among 26 longitudinal lineages, (C) Transition-to-transversion ratio for 26 *Achromobacter* longitudinal lineages. The red horizontal line marks the Ts/Tv ratio of 3.

Substitution rate and hypermutational signatures

Eleven of the 26 longitudinally sampled lineages showed an increased number of SNPs when accounting for the time period between isolates, and we defined these lineages as hypermutable (Figure 3B). The excess number of SNPs in the hypermutable lineages was driven by an excess number of transition substitutions (higher than three-fold transition-to-transversion (Ts/Tv) ratios were identified in hypermutator lineages; Figure 3C). We aimed to identify if any of the hypermutator lineages had insertions, deletions or frameshifts in the mismatch repair (MMR) system and a total of seven mutations were identified in *mutL* (p.Ala376fs; p.Ala384delinsProPro; p.Ala384fs; p.Ala444-Ala445insAlaLeuAlaProGlnAla; p.Thr369-Pro370delinsAlaSerThrAla) and *mutS* (p.Gln174–Ala185del; p.Glu435fs) genes. Of the 61 lineages, 18 (30%) contained at least one of these mutations in genes coding for the MMR system (Figure 4A). Two hypermutable lineages did not have any insertions, deletions, or frameshift mutations in *mutL* and *mutS* genes while two non-hypermutator lineages had mutations in the MMR system genes. In sum, all nine longitudinal *A. ruhlandii* lineages were identified as hypermutators while only one lineage was defined as hypermutator for both *A. insuavis* and *A. xylosoxidans*.

Within-patient bacterial substitution rate was estimated for lineages where three or more *Achromobacter* isolates were available (four lineages for *A. ruhlandii* and *A. xylosoxidans*, and one for *A. insuavis*). The estimated substitution rates for *A. ruhlandii* were on average $4.0 \cdot 10^{-6}$ ($1.66 \cdot 10^{-6}$–$6.23 \cdot 10^{-6}$) SNPs/year/site, for *A. insuavis* substitution rate was estimated to be $1.45 \cdot 10^{-7}$ SNPs/year/site, for *A. xylosoxidans*—on average $3.24 \cdot 10^{-7}$ ($6.48 \cdot 10^{-8}$–$3.48 \cdot 10^{-7}$) SNPs/year/site (one hypermutator lineage with the mutation rate of $2.36 \cdot 10^{-6}$ SNPs/year/site was excluded). These substitution rates correspond to an average of 25.7, 0.7, and 1.9 SNPs/year/genome for *A. ruhlandii*, *A. insuavis*, and *A. xylosoxidans*, respectively (Table S6).

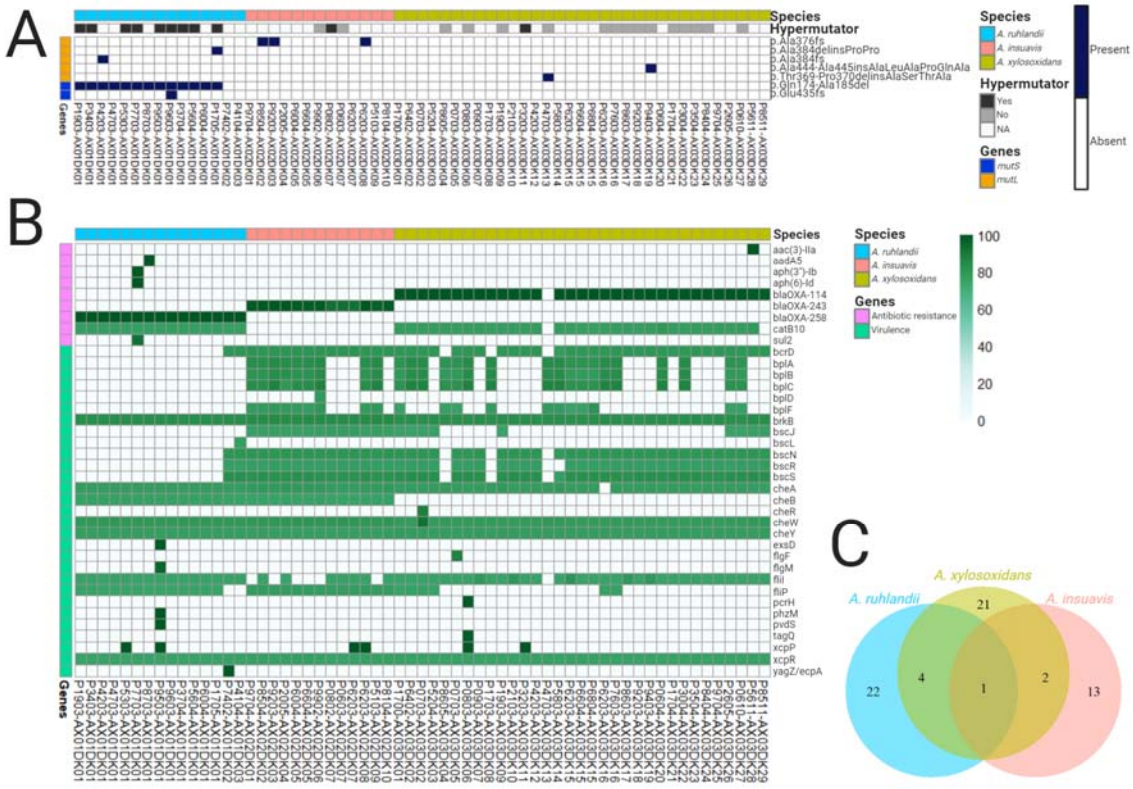

Figure 4. Overview of genetic determinants related hypermutation, virulence and antibiotic resistance, and pathoadaptive genes. (A) Heatmap of insertion, deletion, and frameshift mutations observed in *Achromobacter* lineages in MMR genes. (B) Antibiotic resistance and virulence gene ortholog distribution among lineages. (C) Venn diagram of the most frequently mutated genes and their overlap between the three *Achromobacter* species.

Virulence and antibiotic resistance genes in *Achromobacter*

From *de novo* assembled genomes of the 99 *Achromobacter* isolates we identified virulence (VFDB) and antibiotic resistance (Resfinder) gene orthologs with the Abricate tool. On average, bacterial isolates carried 2 (0–5) antibiotic resistance gene orthologs and 12 (8–18) virulence gene orthologs. The most frequently carried antibiotic resistance gene was the ortholog of *catB10* which codes for chloramphenicol acetyltransferase. Interestingly, the gene was carried by all *A. ruhlandii* and nearly all (31/33) *A. xylosoxidans* lineages but none of the *A. insuavis* isolates (Figure 4B). Furthermore, OXA-type class D β-lactamase $bla_{OXA-258}$ was observed in all *A. ruhlandii* isolates, $bla_{OXA-243}$—in all *A. insuavis* and $bla_{OXA-114}$—in all but one (347-AX03DK13) *A. xylosoxidans* lineages.

Virulence genes *xcpR*, *brkB*, *cheW*, and *cheY* were present in all 61 lineages. Furthermore, the majority of virulence gene orthologs belonged to the type III secretion system (N=8), adhesion (N=5), cell invasion (N=5), endo- (N=5) and exotoxin (N=4) gene orthologs (Figure 4B; Table S7).

Genetic adaptation: Mutation of the same genes across lineages

To explore within-host convergent evolution in *Achromobacter*, we first identified the genes which were most frequently mutated within each species. Genes were defined as frequently mutated if they were among 1% most commonly mutated genes between lineages. If more than 1% of the genes were mutated with the same frequency, those genes were also included in the analysis. A total of 27, 16, and 28 genes were identified as most frequently mutated for *A. ruhlandii*, *A. insuavis*, and *A. xylosoxidans*, respectively (Figure 4C). The clusters of orthologous groups (COG) functional annotations are summarized in Figure 5A for all species (Table S8 for detailed information) with the highest mutation frequency in genes coding for signal transduction; inorganic ion transport and metabolism; replication, recombination and repair; and transcription.

Although knowledge is lacking about many *Achromobacter* gene functions, bacterial sequence similarity analysis allowed us to identify possible antibiotic resistance and virulence-related genes among the most frequently mutated genes in *Achromobacter*. After manual literature search 10, 4 and 4 genes were defined as related to antibiotic resistance in *A. ruhlandii*, *A. insuavis* and *A. xylosoxidans*, respectively, whereas 8, 2 and 7 genes were defined as virulence-related genes (Additional File 1: Table S9).

Table 2. Seven most frequently mutated genes and their function.

RefSeq ID	No. of species	No. of lineages	Product	Function
WP_006389199.1	3	10	DNA-binding transcriptional regulator AxyZ	Antibiotic resistance
WP_006227290.1	2	5	NAD(P)/FAD-dependent oxidoreductase	Metabolic pathways
WP_013391523.1	2	10	penicillin-binding protein 2	Antibiotic resistance
WP_006392856.1	2	8	ABC transporter substrate-binding protein	Transport
WP_006394390.1	2	8	multidrug efflux RND transporter permease	Antibiotic resistance
WP_006387572.1	2	8	Signal transduction histidine kinase	Two-component signaling
WP_006221301.1	2	9	TonB-dependent hemin, ferrichrome receptor	Transport

One gene was frequently mutated in all three species and six genes were observed as most frequently mutated in two of the *Achromobacter* species (Table 2). Ortholog search of a previously defined list of 52 pathoadaptive *P. aeruginosa* genes revealed three orthologs among the most frequently mutated *Achromobacter* genes: *mexZ* (WP_006389199.1), *mexB* (WP_024068614.1) and *gyrA* (WP_049072335.1). The most notable convergent evolution was observed for the transcriptional regulator *axyZ* (*mexZ* ortholog) gene which was frequently mutated in all 3 species, i.e. 10 out of 26 lineages acquiring a mutation over the infection period.

Finally, the ratio of non-synonymous to synonymous substitutions (dN/dS) was significantly different between the 1% most frequently mutated genes and non-frequently mutated genes (Figure 5C) in *A. ruhlandii* and *A. xylosoxidans* (Fisher's exact test; $p<2.2\cdot10^{-16}$ and $p=1.6\cdot10^{-6}$, respectively).

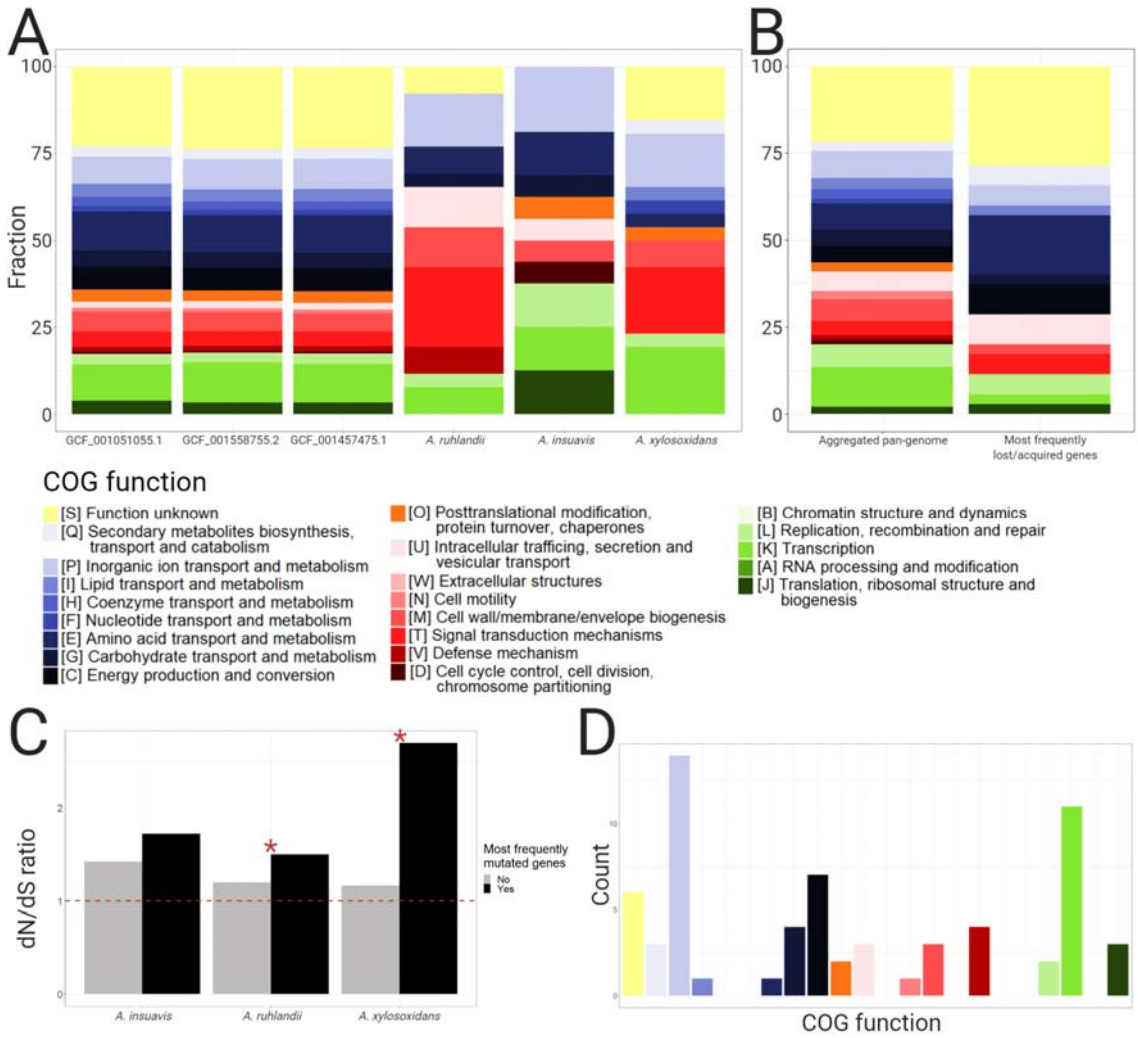

Figure 5. Functional groups and relative rate of nonsynonymous mutations of genes associated with host adaptation. COG annotation of (A) three *Achromobacter* reference genomes and the most frequently mutated genes, and (B) the aggregated pan-genome and most frequently lost or acquired genes. (C) Non-synonymous to synonymous (dN/dS) mutation ratio between most frequently mutated and non-frequently mutated genes in each *Achromobacter* species. Red stars mark significant differences between two groups after performing Fisher's exact test. (D) COG annotations of the most significant unitigs associated with six different antibiotic resistance phenotypes.

Gene presence-absence analysis

As a crucial part of bacterial adaptation and evolution, we investigated the most frequently lost and acquired genes in the 26 longitudinally sampled *Achromobacter* lineages. We defined a gene as lost if it was present in the first isolate but absent in one or more of the later

isolates and a gene was defined as acquired if it was absent in the first isolate but present in one or more of the later isolates. The lineage pan-genomes contained on average 5,940 (5,709–6,377) genes of which on average 62 (14–183) were lost or acquired. In total, we found 1,464 genes to be variable within lineages and 735 of these were unique to a single lineage in the aggregated pan-genome (Figure 3A). We observed that genes were 1.8 times more often lost than acquired and lost/acquired in groups 37 times more commonly than individually (Table S10).

Most frequently lost or acquired genes were defined if a gene was lost/acquired in a minimum of 3 lineages. In total, 38 genes passed these criteria; however, after the manual inspection, one gene was defined as a false positive: the gene was manually assessed as present, yet the gene was not assembled with a minimum requirement of 25% of the gene length coverage, thus was predicted to be absent by GenAPI. The remaining 37 genes that were frequently lost or acquired, were annotated with EGGNOG-mapper: transport and metabolism genes (Figure 5B, colored in blue) were more frequently lost/acquired, especially amino acid transport and metabolism genes, when compared to the composition of the aggregated pan-genome. Contrary, structural, cellular organization and cell cycle genes; reproduction; and gene expression regulating genes were less frequently lost/acquired (Figure 5B). Overall, aggregated pan-genome gene composition is comparable to individual reference genome gene composition (Figure 5A and 5B).

Genome-wide association between *Achromobacter* genotypes and antibiotic resistance

To test for associations between bacterial genetics and phenotypic antibiotic resistance we performed unitig-based DBGWAS analysis. Out of 21 antibiotics where the resistance was phenotypically tested, only 10 had both susceptible and resistant isolates from all three *Achromobacter* species for which the association analysis was performed (see Materials and

Methods for detailed information; Figure S2; Table S11). No unitigs passed a 5% FDR corrected q-value threshold for AMC, CAZ, CST, and TGC. Ten most significant unitigs were used for the six remainder association test analysis, resulting in 60 gene features (49 unique genes) significantly associated with antibiotic resistance phenotypes (Figure 5D; Table S12). The majority of associated unitigs were annotated by EGGNOG-mapper as belonging to transcription regulation (N=11) or inorganic ion transport (N=14) genes. Furthermore, energy production and conversion (N=7), carbohydrate production and conversion (N=4), and defense mechanism (N=4) genes were also observed to be significantly associated with the resistance phenotype in one or more antibiotics. Of 49 genes, 12 were associated with known antibiotic resistance phenotypes and were previously described as important for antibiotic resistance development in other bacteria (Table S12).

Interestingly, 7 out of 29 *A. ruhlandii* (6 out of 27 DES) isolates were resistant or intermediately resistant to all 21 antibiotics while only one *A. insuavis* and all of the *A. xylosoxidans* isolates were not susceptible to any antibiotic.

Discussion

Achromobacter is an emerging pathogen causing chronic respiratory tract infections in patients with CF; however, the genetic epidemiology of these infections is little understood. We sequenced and analyzed 101 genomes of *Achromobacter* isolates from 51 patients with CF which is the largest longitudinally collected *Achromobacter* genome dataset available to-date. This allowed us to investigate the population genomics, within-host adaptation, mutational mechanisms, including antibiotic resistance phenotype and bacterial genetics association, and patient-to-patient transmission events with high sensitivity. Furthermore, we defined a comprehensive analytical framework that can be applied to other research involving exploration of within-host adaptation of pathogenic bacteria (Figure 6).

Our genomic analysis showed that MALDI-TOF/API 20 is not accurate for *Achromobacter* species-level typing as it has also been recently indicated by others (18,19); thus we suggest that sequencing of marker genes (e.g., bla_{OXA} or *nrdA*) or whole genome (WGS) should be used for species typing in the clinical setup.

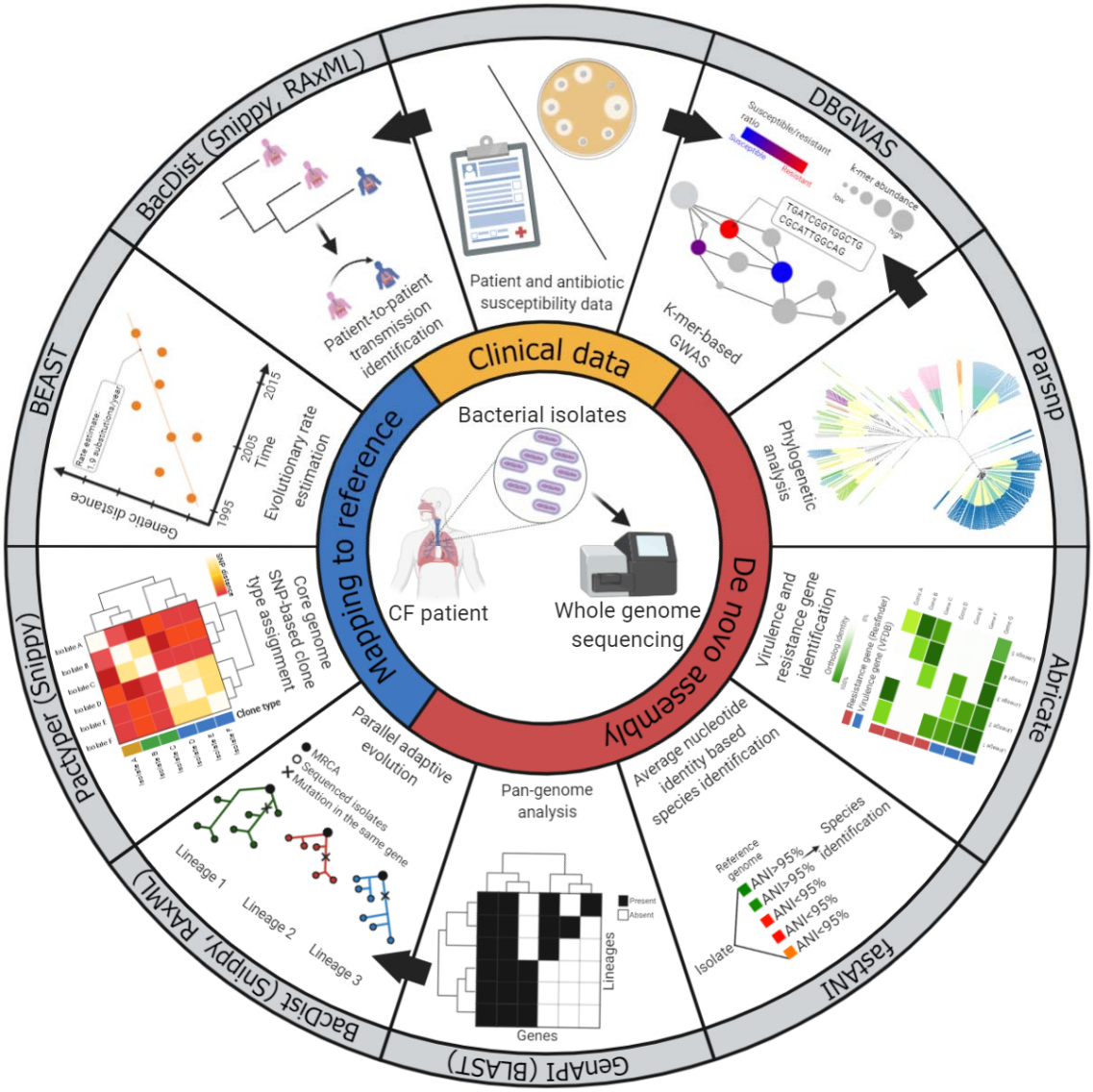

Figure 6. Analytical framework used for *Achromobacter* genome exploration in this study.

Core genome sizes were comparable between *A. ruhlandii*, *A. insuavis* and *A. xylosoxidans* even though the number of accessory and unique genes in species' pan-genomes varied greatly. *A. ruhlandii* isolates mostly belong to DES (as confirmed by the phylogenetic analysis) which are known to be spread through patient-to-patient transmission and, therefore, have lower pan-genome plasticity than *A. xylosoxidans* or *A. insuavis*. Moreover, the phylogenetic analysis from patients attending Aarhus and Copenhagen CF centers also suggests multiple transmission events between patients attending different CF centers. Furthermore, recent first-time isolation of DES in two patients with CF (P7703 and P8703) indicates that the transmission of this strain might not be under complete control and requires additional actions to be taken for the eradication of DES. Nearly 20% of patients were infected by multiple *Achromobacter* species or clone types over the sampled time-period which could indicate that not all *Achromobacter* caused airway infections lead to chronic infections and further supports the early antibiotic treatment of *Achromobacter* infections. (10)

A comparison of the publicly available *Achromobacter* genomes from the RefSeq database revealed that our dataset well represented the genetic diversity of *A. xylosoxidans* and *A. insuavis* species. However, because of DES overrepresentation, *A. ruhlandii* isolates did not reflect the species genetic diversity. Furthermore, ANI and core-genome based phylogenetic analysis revealed that more than 10% of publicly available *Achromobacter* genomes are supposedly misannotated in the RefSeq database which we anticipated to unravel to ease future research on *Achromobacter*.

The majority of *Achromobacter* infections are acquired from the environment and prevalence of patient-to-patient transmission remains controversial: while some studies identified patients being infected with unique clone types of *Achromobacter* (20–22) other studies

reported the cases of suspected patient-to-patient transmission. (23,24) We identified multiple transmission events in each *Achromobacter* species which occasionally were transmitted to multiple patients. However, unlike in *P. aeruginosa*, the observed SNP distance between the transmitted isolates was higher (13); hence, complicated the successful transmission identification and might explain why no patient-to-patient transmission was discovered in previous studies. (20–22) Interestingly, DES—a known hypermutator—is widely transmitted between patients with CF even though some studies show reduced transmissibility of hypermutator strains. (25) Finally, we observed that infection of the same clone type in several patients was always a sign of patient-to-patient transmission of *Achromobacter*. This knowledge could be applied in diagnostics where sharing an *Achromobacter* clone type would be treated as the first sign of suspected transmission between patients and would lead to further testing.

Unlike many bacterial species, substitution rates for *Achromobacter* are not known. (26) Here, we successfully estimated the within-host substitution rate for *A. xylosoxidans* species included in the analysis. Our *A. ruhlandii* dataset only consisted of hypermutator lineages which led to a high substitution rate estimate, and our substitution rate estimate for *A. insuavis* was based on only one longitudinally sampled lineage. *A. xylosoxidans* ($3.24 \cdot 10^{-7}$ SNPs/year per site) substitution rate is comparable to other Gram-negative bacterial species: *P. aeruginosa* ($4.0 \cdot 10^{-7}$ SNPs/year per site) (27), *Shigella sonnei* ($6.0 \cdot 10^{-7}$ SNPs/year per site) (28), *Vibrio cholerae* ($8.0 \cdot 10^{-7}$ SNPs/year per site) (29) and *Yersinia pestis* ($2.0 \cdot 10^{-8}$ SNPs/year per site) (30). Moreover, our identified mutations in MMR genes could explain the majority of the observed hypermutable lineages. One recently defined DES *mutS* deletion (p.Gln174–Ala185del) (11) was also observed in all our DES isolates supporting the hypothesis of this deletion causing DES hypermutability.

Moreover, ortholog virulence gene analysis revealed that *Achromobacter* carries several virulence factors without clear differences between species with several virulence gene orthologs coding for host cell invasion (*cheW* and *cheY*) and facilitating evasion of the host immune response (*brkB* (31)) were observed in all lineages. Furthermore, contrary to being commonly found among *A. insuavis* and *A. xylosoxidans*, 9 virulence gene orthologs (*bcrD*, *bplA*, *bplB*, *bplC*, *bplF*, *bscJ*, *bscN*, *bscR*, and *bscS*) were not identified in any of the DES lineages, which are well-adapted to human airway, further supporting the adaptive trade-off evolution hypothesis that virulence genes are not required or are selected against in chronic infections. (26)

While it was previously suggested to use bla_{OXA} genes for *Achromobacter* species typing (32–34), we showed that in some cases such strategy would not be sufficient for species identification as isolates can carry none of the bla_{OXA} genes. Furthermore, none of the *A. insuavis* isolates carried *catB10* chloramphenicol resistance gene; however, these observations alone were not sufficient to explain the differences in antibiotic resistance phenotype between lineages and species.

Among the candidate pathoadaptive genes (frequently mutated and lost/acquired genes), we identified multiple antibiotic resistance genes which were markedly more frequently mutated among *A. ruhlandii* isolates than *A. insuavis* or *A. xylosoxidans* isolates. This phenomenon might signal about the continuous adaptive evolution even in highly antibiotic-resistant strains. (7) Nonetheless, more antibiotic resistance and virulence genes among frequently mutated genes might be identified if gene annotation of *Achromobacter* reference genomes improved. Furthermore, *axyZ* (*mexZ* ortholog), which was the only frequently mutated gene in all three lineages, is involved in the development of multidrug resistance by regulating AxyXY-OprZ RND-type efflux system, hence is crucial during adaptation to the host

environment. (35) In sum, our identified candidate pathoadaptive genes (belonging to defense mechanism; signal transduction; inorganic ion transport and metabolism; transcription and replication gene functional classes) are comparable to the observations in *P. aeruginosa* infecting patients with CF (13) and other smaller-scale studies on *Achromobacter*. (16) The observed gene loss and acquisition patterns were comparable to the ones observed in *P. aeruginosa*: genes were almost 2 times more often lost than acquired and 37 times more frequently lost in groups than individually; however, unlike in *P. aeruginosa*, no convergent loss or acquisition of gene clusters was observed. (36)

High between-patient diversity of *Achromobacter* is comparable to previous observations of Marvig *et al.* (2015) (13) in *P. aeruginosa*. Noticeably higher dN/dS ratios among the most frequently mutated genes when compared with non-frequently mutated genes indicate adaptation and selective pressure from the environment for these genes to be mutated. Additionally, high dN/dS ratios in both frequently and non-frequently mutated genes show that more than top 1% mutated genes are under selective pressure; nonetheless, a larger dataset is needed to identify more genes without sacrificing analysis accuracy.

To further explore the differences in antibiotic susceptibility between *Achromobacter* isolates, we performed a k-mer based GWAS analysis. Limited number of bacterial isolates and high innate resistance to certain antibiotics restricted our analysis to only 6 successful GWAS associations which revealed that transcription regulation and inorganic ion transport genes are likely playing a role in *Achromobacter* antibiotic resistance development. These findings are consistent with our identified putative pathoadaptive genes indicating that presence of antibiotics might be driving the selective pressure. Inorganic ion transport gene changes could have a secondary influence on antibiotic resistance as such changes help

overcome the problem of iron deficiency in the human airways allowing better intrinsically resistant bacteria survival despite the presence of antibiotics. Similar patterns were previously identified in other bacteria causing chronic infections in patients with CF. (26,37,38) Another frequently associated feature was transcription regulation genes; changes in such genes might modify the transcription of intrinsic antibiotic resistance genes. This phenomenon was previously described in both *Achromobacter (39)* and other bacterial pathogens (40,41). Nevertheless, including more isolates for the bacterial GWAS analysis could reveal more putative genomic features associated with antibiotic resistance phenotype.

Our study has several limitations. First, even larger studies are necessary to further characterize and identify the genetic epidemiology of *Achromobacter* infections. Second, the lack of genome annotation and overall knowledge about *Achromobacter* limited the interpretation of putative pathoadaptive genes and genes associated with resistance phenotypes. Finally, a single isolate at a given time point is not sufficient to completely reflect the genetic diversity of the bacterial population (42,43); therefore, some of our findings might be the result of diversification and not the fixation of the adaptive mutations in *Achromobacter*.

In conclusion, by using the largest dataset to-date of *Achromobacter* clinical isolates from patients with CF, we used a comprehensive analytical framework for thorough bacterial genomic data exploration which could facilitate future pathogenic bacteria genomic studies. Our work revealed that WGS can be used for successful *Achromobacter* typing and patient-to-patient transmission identification. Our analytical framework allowed us to explore the *Achromobacter* mutational and transmission patterns, identify putative pathoadaptive genes, locate virulence and antibiotic resistance genes. The gained knowledge allows us to better

understand the requirements for successful *Achromobacter* adaptation during chronic infection in airways of patients with CF which could help predict clinical progression of *Achromobacter* infections, further the development of urgently needed better treatment strategies and patient-to-patient transmission prevention.

Materials and methods

Bacterial isolates

The analysis included 101 clinical isolates of *Achromobacter* that prior to this study were identified in the routine clinical microbiology laboratory as *A. xylosoxidans* by API N20 (bioMérieux) or MALDI-TOF typing (Bruker, Germany). The isolates were sampled from 51 patients with CF attending the Copenhagen Cystic Fibrosis Center at Rigshospitalet, Denmark. Over the timespan of 0–18 years (median 6.5 years), 64 isolates from 25 patients, belonging to the same species and clone type, were longitudinally collected (median 2 isolates) and 37 isolates from 29 patients were single isolates.

Bacterial genome sequencing and definition of clone type

Genomic DNA was prepared from *Achromobacter* isolates on a QIAcube system using a DNeasy Blood and Tissue kit (Qiagen) and sequenced on an Illumina HiSeq 2000 platform, generating 250-bp paired-end reads and using a multiplexed protocol to obtain an average of 1,124,551 reads (range of 350,677–2,118,817) for each of the genomic libraries. Clone types were defined by Pactyper (https://github.com/MigleSur/Pactyper) (44) using the default parameters and species' core genome defined by GenAPI. (45,46) Lineage was defined by all isolates belonging to the same species and the same clone type.

Bacterial genome assembly

Sequence reads from each isolate were corrected and assembled by SPAdes version 3.10.1 (47) using default parameters and k-mer sizes ranging from 21 to 127. Assembled contigs were joined to 216 scaffolds on average (92–506) and were used for further analysis.

***De novo* assembly-based phylogenetic tree generation**

Core genome SNP-based phylogenetic trees of 101 *de novo* assembled *Achromobacter* isolates and DES isolates were generated using parsnp version 1.2 (48) with the default settings and without a reference genome. Eight complete *Achromobacter* reference genomes were included in the phylogenetic analysis of 101 *de novo* assembled isolates (RefSeq assembly accession: GCF_000165835.1, GCF_000758265.1, GCF_001051055.1, GCF_001457475.1, GCF_001558755.2, GCF_001558915.1, GCF_001559195.1 and GCF_900475575.1). The phylogenetic tree of Copenhagen and Aarhus DES isolates was visualized using iTOL webtool using the earliest available isolate DES28 as a root. (49)

The joint phylogenetic analysis of all publicly available *Achromobacter* genomes from RefSeq database (50) (access date: 2019.10.29) and clinical *Achromobacter* isolates from patients with CF was performed on core genome SNPs by RAxML (51) using GTRCAT model. The core genome was defined during the aggregated pan-genome construction and was used as a reference for variant calling from *de novo* assemblies with Snippy. (52) Phylogenetic trees were visualized with Microreact webservice. (53)

Average nucleotide identity calculation

Wrong public *Achromobacter* genome annotation identification from RefSeq database (50) and separation of CF isolates of different species was performed by calculating ANI with fastANI version 1.11. (54) 95% threshold was used to define two bacterial isolates as different species.

Aggregated pan-genome generation, characterization and visualization

An aggregated pan-genome was created by clustering all pan-genomes from longitudinal lineages and *de novo* assemblies from single-isolate lineages with GenAPI. (46) Every gene in the aggregated pan-genome was then aligned back to the individual pan-genomes/*de novo* assemblies to determine if the gene is (1) non-present in the lineage, (2) present and variable within the lineage or (3) present and non-variable. A matrix for an aggregated pan-genome was generated for 26 longitudinal lineages and 35 single-isolate lineages, and visualized using R version 3.3.3 (55) with pheatmap library version 1.0.8. (56)

Bacterial genome alignment and variant calling

Alignments, variant calling and pairwise SNP distance identification for *Achromobacter* isolates were performed using reference genome (GCF_001051055.1 for *A. ruhlandii* (AX01 group), GCF_001558755.2 for *A. insuavis* (AX02 group) and GCF_001457475.1 for *A. xylosoxidans* (AX03 group)) with BacDist (https://github.com/MigleSur/BacDist) (57) workflow that is based on variant calling with Snippy. (52) Sequence alignments on average included 84% (81.72%–89.58%) of the raw sequencing reads for *A. ruhlandii*, 87% (75.45%–92.57%) for *A. insuavis* and 86% (75.95%–93.61%) for *A. xylosoxidans*. Low-quality variants or variants shared among all isolates were discarded by BacDist. Raw Snippy variant calls were used for the identification of variants in the DNA mismatch repair system (MMR). SNP distances and phylogenetic analysis from BacDist were used for patient-to-patient transmission identification.

Hypermutator identification

Hypermutators were identified in longitudinal lineages by evaluating the number of pairwise SNP distances between isolates over the time interval (BacDist analysis (57)). If 10 or more SNPs/year were on average observed to be introduced, the lineage was concluded to be

hypermutable. Insertions, deletions and frameshifts in the MMR system *mutL* and *mutS* genes were evaluated to identify which genetic changes could cause hypermutability. (58)

Substitution rate and MRCA estimation

The bacterial substitution rate (59) estimation was performed for each lineage containing 3 or more isolates (11 lineages in total) by using BEAUti (BEAST version 2.5.0). (60) MRCA was attempted to be estimated for all isolates where patient-to-patient transmission was suspected. Sequence alignments from BacDist were used as input with the following parameters: (1) sequences were annotated with the sampling date ("dated tips"), (2) GTR substitution model, (3) uniform prior for clock rate, (4) prior for population size: 1/X, (5) tree prior: coalescent constant population, (6) Gamma priors for substitution. MCMC was run for 50,000,000 iterations. Convergence was checked by inspecting an effective sample size and parameter value traces in the Tracer software (version 1.7.1). (61) Multiple tests for each sample were performed to ensure reproducibility and convergence. The obtained clock rate (per site per year) was multiplied by the alignment size to obtain a substitution rate per genome per year.

Virulence and antibiotic resistance gene identification

Orthologs of resistance and virulence genes in 61 *Achrmobacter* lineages were identified by Abricate tool (62) using *de novo* assemblies of *Achromobacter* isolates and the corresponding databases (VFDB (containing 2597 genes; retrieved: 2018.03.21) (63) for virulence genes and Resfinder (containing 2280 genes; retrieved: 2018.03.21) (64) for resistance genes). Gene ortholog was considered present in the corresponding database if the alignment made up minimum 50% of the gene length and its identity was minimum 75%.

Frequently mutated gene definition

Most frequently mutated genes were defined as the top 1% of all mutated genes for the species. If there were more genes mutated with the same frequency as the 1% most frequently

mutated genes, these genes were also included in the final analysis. The identified most mutated genes were annotated by EGGNOG-mapper version 1.0.3 (65) using DIAMOND, EGGNOG's bacterial database, $1 \cdot 10^{-8}$ e-value cutoff, and 0.8 minimum query sequence coverage settings. The predicted COGs were used for further gene analysis.

Pseudomonas aeruginosa orthologs were identified by performing clustering with CD-HIT (66) using word size of 3 and 50% identity thresholds. Joint *Achromobacter* most frequently mutated genes were identified by clustering with CD-HIT (66) with word size of 3 and 80% identity threshold.

Gene loss/acquisition analysis

De novo assembled genomes were annotated using Prokka version 1.12 (67) with the settings of a minimum contig length of 200 nucleotides and using a manually created annotation database for *Achromobacter* (GCF_001051055.1 (*A. ruhlandii*), GCF_001558755.2 (*A. insuavis*) and GCF_001457475.1) (*A. xylosoxidans*). Further analysis was based on longitudinal lineages by GenAPI software with default settings. (45) Genes shorter than 150 nucleotides were excluded from further analysis by default. Gene presence/absence matrices were visualized by using R version 3.3.3 (55) and pheatmap library version 1.0.8. (56)

Antibiotic resistance-associated gene identification

DBGWAS version 0.5.4 (68) software was used for bacterial genome-wide association analysis using 10 (Amoxicillin-Clavulanate (AMC), Ceftazidime (CAZ), Chloramphenicol (CHL), Colistin (CST), Imipenem (IPM), Meropenem (MEM), Piperacillin-Tazobactam (TZP), Sulfamethizole (SMZ), Tigecycline (TGC) and Trimethoprim-Sulfamethoxazole (SXT)) different antibiotic resistance phenotypes and *de novo* assembled scaffolds of 92 isolates for which the antibiotic susceptibility profiles were available. Core genome SNP-based phylogenetic tree was used to correct for population structure while all available annotations of *Achromobacter* genes from UniProt database (69) were used for unitig

annotation (271,851 genes; retrieved: 2020.04.19). Ten most significant unitigs for each antibiotic test were used for further analysis as the tool authors advise against using a p-value threshold when testing several phenotypes. (68)

List of abbreviations

AMC: Amoxicillin-Clavulanate

ANI: Average nucleotide identity

API N20: Analytical profile index N20

CAZ: Ceftazidime

CHL: Chloramphenicol

COG: Clusters of Orthologous Groups

CST: Colistin

CF: Cystic fibrosis

DES: Danish epidemic strain

dN/dS: Non-synonymous to synonymous ratio

GWAS: Genome-wide association study

IPM: Imipenem

MALDI-TOF: Matrix assisted laser desorption ionization-time of flight mass spectrometry

MEM: Meropenem

MMR: Mismatch repair

MRCA: Most recent common ancestor

TZP: Piperacillin-Tazobactam

SMZ: Sulfamethizole

SNP: Single nucleotide polymorphism

TGC: Tigecycline

Ts/Tv: Transition-to-transversion ratio

SXT: Trimethoprim-Sulfamethoxazole

WGS: Whole genome sequence

Availability of data and materials

Achromobacter whole genome sequencing data is available at European Nucleotide Archive under study accession number PRJEB39108.

Acknowledgments

Ulla Johansen is thanked for expert technical assistance and Niels Høiby is thanked for collecting the earliest *Achromobacter* isolates. All figures were partly or completely created using BioRender (https://biorender.com/).

Funding

This work was supported by the Danish Cystic Fibrosis Association (Cystisk Fibrose Foreningen) and the Danish National Research Foundation (grant number 126). HKJ was supported by The Novo Nordisk Foundation as a clinical research stipend (NNF12OC1015920), by Rigshospitalets Rammebevilling 2015-17 (R88-A3537), by Lundbeckfonden (R167-2013-15229), by Novo Nordisk Fonden (NNF15OC0017444), by RegionH Rammebevilling (R144-A5287) by Independent Research Fund Denmark / Medical and Health Sciences (FTP-4183-00051) and by 'Savværksejer Jeppe Juhl og Hustru Ovita Juhls mindelegat'.

Disclosure declaration

We declare no conflict of interest.

Contributions

R.L.M. and H.K.J conceived the study. R.L.M., H.K.J. and F.C.N supervised the study. M.G., R.L.M. and H.K.J. defined the methodology. M.G. and R.L.M. designed the bioinformatics

workflows for the analysis. M.G. conducted the analysis. M.G., N.N.L. and R.L.M. analyzed and interpreted the data. M.G. prepared the manuscript draft and visualizations. R.L.M., H.K.J., F.C.N and N.N.L. reviewed and edited the draft.

Ethics declarations

Use of the stored clinical isolates was approved by the local ethics committee at the Capital Region of Denmark RegionH (registration number H-4-2015-FSP).

Supplementary information

Table S1. ANI between eight complete *Achromobacter* genome RefSeq sequences used in this study.

Table S2. Table of publicly available RefSeq *Achromobacter* sequences with original and new proposed annotations which were based on ANI and core genome SNP-based phylogeny. Colored cells in the second column corresponds to the annotation which does not match with the new annotation. Isolates belonging to the same undescribed species are colored in the third column with the same color.

Table S3. Estimated date of patient-to-patient transmission of *Achromobacter* isolates based on BEAST analysis and inferred time since most recent common ancestor (MRCA).

Table S4. Smallest SNP distances within and between lineages involved in patient-to-patient transmission.

Table S5. Number of SNPs in early and late isolates in the 26 *Achromobacter* lineages.

Table S6. Estimated clock rate, mutation rate per genome per year, tree height and time interval between earliest and latest isolates in *Achromobacter* lineages where mutation rate was estimated.

Table S7. A list of ortholog virulence genes present in *Achromobacter* and their function based on VFDB gene annotations.

Table S8. List of most frequently mutated genes in *Achromobacter* with their COG function, corresponding E-value and gene annotation.

Table S9. List of most frequently mutated genes in *A. ruhlandii*, *A. insuavis* and *A. xylosoxidans* with gene description and gene relatedness to antibiotic resistance or virulence. References defining gene's relatedness to antibiotic resistance or virulence are provided where relevant.

Table S10. Table of pan-genome, core, accessory and unique genes in *Achromobacter* lineages. Variable genes are separated by whether they were (1) lost or acquired and (2) variable in a group or individually.

Table S11. Isolate susceptibility data to 21 tested antibiotics. R - resistant to the antibiotic, I - intermediately resistant to the antibiotic, S - susceptible to the antibiotic.

Table S12. Ten most significant GWAS associations (q-value<0.05) for each antibiotic susceptibility association test.

Figure S1. Core genome SNP phylogenetic trees of suspected bacterial isolates transmission between patients in A. ruhlandii, A. insuavis, and A. xylosoxidans where 4 or more isolates were available. Isolates are named as follows: [Patient ID] [Sampling date] [Species and clone type].

Figure S2. Summary of A. ruhlandii, A. insuavis and A. xylosoxidans susceptibility profiles to 21 tested antibiotics.

References

1. Martin LW, Robson CL, Watts AM, Gray AR, Wainwright CE, Bell SC, et al. Expression of Pseudomonas aeruginosa Antibiotic Resistance Genes Varies Greatly during Infections in Cystic Fibrosis Patients. Antimicrob Agents Chemother. 2018 Oct 24;62(11).
2. Folkesson A, Jelsbak L, Yang L, Johansen HK, Ciofu O, Høiby N, et al. Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective. Nat Rev Microbiol. 2012 Dec;10(12):841–51.

3. Baishya J, Wakeman CA. Selective pressures during chronic infection drive microbial competition and cooperation. npj Biofilms and Microbiomes. 2019 Jun 7;5:16.

4. McGuigan L, Callaghan M. The evolving dynamics of the microbial community in the cystic fibrosis lung. Environ Microbiol. 2015 Jan;17(1):16–28.

5. Jorth P, Ehsan Z, Rezayat A, Caldwell E, Pope C, Brewington JJ, et al. Direct Lung Sampling Indicates That Established Pathogens Dominate Early Infections in Children with Cystic Fibrosis. Cell Rep. 2019 Apr 23;27(4):1190-1204.e3.

6. Callaghan M, McClean S. Bacterial host interactions in cystic fibrosis. Curr Opin Microbiol. 2012 Feb;15(1):71–7.

7. Ridderberg W, Nielsen SM, Nørskov-Lauritsen N. Genetic Adaptation of Achromobacter sp. during Persistence in the Lungs of Cystic Fibrosis Patients. PLoS ONE. 2015 Aug 27;10(8):e0136790.

8. Stobbelaar K, Van Hoorenbeeck K, Lequesne M, De Dooy J, Ho E, Vlieghe E, et al. Sepsis Caused by Achromobacter Xylosoxidans in a Child with Cystic Fibrosis and Severe Lung Disease. Am J Case Rep. 2016 Aug 8;17:562–6.

9. Mahenthiralingam E. Emerging cystic fibrosis pathogens and the microbiome. Paediatr Respir Rev. 2014 Jun;15 Suppl 1:13–5.

10. Wang M, Ridderberg W, Hansen CR, Høiby N, Jensen-Fangel S, Olesen HV, et al. Early treatment with inhaled antibiotics postpones next occurrence of Achromobacter in cystic fibrosis. J Cyst Fibros. 2013 Dec;12(6):638–43.

11. Ridderberg W, Jensen Handberg K, Nørskov-Lauritsen N. Prevalence of hypermutator isolates of Achromobacter spp. from cystic fibrosis patients. Int J Med Microbiol. 2020 Jan 16;151393.

12. Gade SS, Nørskov-Lauritsen N, Ridderberg W. Prevalence and species distribution of Achromobacter sp. cultured from cystic fibrosis patients attending the Aarhus centre in Denmark. J Med Microbiol. 2017 May 8;66(5):686–9.

13. Marvig RL, Sommer LM, Molin S, Johansen HK. Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. Nat Genet. 2015 Jan;47(1):57–64.

14. Brockhurst MA, Harrison E, Hall JPJ, Richards T, McNally A, MacLean C. The ecology and evolution of pangenomes. Curr Biol. 2019 Oct 21;29(20):R1094–103.

15. Yang L, Jelsbak L, Marvig RL, Damkiær S, Workman CT, Rau MH, et al. Evolutionary dynamics of bacteria in a human host environment. Proc Natl Acad Sci USA. 2011 May 3;108(18):7481–6.

16. Jeukens J, Freschi L, Vincent AT, Emond-Rheault J-G, Kukavica-Ibrulj I, Charette SJ, et al. A Pan-Genomic Approach to Understand the Basis of Host Adaptation in Achromobacter. Genome Biol Evol. 2017 Apr 1;9(4):1030–46.

17. Swenson CE, Sadikot RT. Achromobacter respiratory infections. Ann Am Thorac Soc. 2015 Feb;12(2):252–8.

18. Garrigos T, Neuwirth C, Chapuis A, Bador J, Amoureux L. Development of a database for the rapid and accurate routine identification of Achromobacter species by matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS). Clin Microbiol Infect. 2020 Apr 10;

19. Papalia M, Steffanowski C, Traglia G, Almuzara M, Martina P, Galanternik L, et al. Diversity of Achromobacter species recovered from patients with cystic fibrosis, in Argentina. Rev Argent Microbiol. 2019 Jun 25;

20. Dunne WM, Maisch S. Epidemiological investigation of infections due to Alcaligenes species in children and patients with cystic fibrosis: use of repetitive-element-sequence polymerase chain reaction. Clin Infect Dis. 1995 Apr;20(4):836–41.

21. Edwards BD, Greysson-Wong J, Somayaji R, Waddell B, Whelan FJ, Storey DG, et al. Prevalence and Outcomes of Achromobacter Species Infections in Adults with Cystic Fibrosis: a North American Cohort Study. J Clin Microbiol. 2017 Apr 26;55(7):2074–85.

22. Vu-Thien H, Moissenet D, Valcin M, Dulot C, Tournier G, Garbarg-Chenon A. Molecular epidemiology of Burkholderia cepacia, Stenotrophomonas maltophilia, and Alcaligenes xylosoxidans in a cystic fibrosis center. Eur J Clin Microbiol Infect Dis. 1996 Nov;15(11):876–9.

23. Cools P, Ho E, Vranckx K, Schelstraete P, Wurth B, Franckx H, et al. Epidemic Achromobacter xylosoxidans strain among Belgian cystic fibrosis patients and review of literature. BMC Microbiol. 2016 Jun 24;16(1):122.

24. Hansen CR, Pressler T, Ridderberg W, Johansen HK, Skov M. Achromobacter species in cystic fibrosis: cross-infection caused by indirect patient-to-patient contact. J Cyst Fibros. 2013 Dec;12(6):609–15.

25. Maciá MD, Blanquer D, Togores B, Sauleda J, Pérez JL, Oliver A. Hypermutation is a key factor in development of multiple-antimicrobial resistance in Pseudomonas aeruginosa strains causing chronic lung infections. Antimicrob Agents Chemother. 2005 Aug;49(8):3382–6.

26. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. Nat Rev Microbiol. 2016 Mar;14(3):150–62.

27. Marvig RL, Johansen HK, Molin S, Jelsbak L. Genome analysis of a transmissible lineage of pseudomonas aeruginosa reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. PLoS Genet. 2013 Sep 5;9(9):e1003741.

28. Holt KE, Baker S, Weill F-X, Holmes EC, Kitchen A, Yu J, et al. Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. Nat Genet. 2012 Sep;44(9):1056–9.

29. Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, et al. Evidence for several waves of global transmission in the seventh cholera pandemic. Nature. 2011 Aug 24;477(7365):462–5.

30. Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, Wagner DM, et al. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nat Genet. 2010 Dec;42(12):1140–3.

31. Fernandez RC, Weiss AA. Cloning and sequencing of a Bordetella pertussis serum resistance locus. Infect Immun. 1994 Nov;62(11):4727–38.

32. Turton JF, Mustafa N, Shah J, Hampton CV, Pike R, Kenna DT. Identification of Achromobacter xylosoxidans by detection of the bla(OXA-114-like) gene intrinsic in this species. Diagn Microbiol Infect Dis. 2011 Jul;70(3):408–11.

33. Papalia M, Almuzara M, Cejas D, Traglia G, Ramírez MS, Galanternik L, et al. OXA-258 from Achromobacter ruhlandii: a species-specific marker. J Clin Microbiol. 2013 May;51(5):1602–5.

34. Traglia G, Papalia M, Almuzara M, Gutkind G, Centrón D, Vay C, et al. Presence of OXA-type enzymes in Achromobacter insuavis and A. dolens. Curr Microbiol. 2014 Oct;69(4):501–6.

35. Bador J, Neuwirth C, Grangier N, Muniz M, Germé L, Bonnet J, et al. Role of AxyZ Transcriptional Regulator in Overproduction of AxyXY-OprZ Multidrug Efflux System in Achromobacter Species Mutants Selected by Tobramycin. Antimicrob Agents Chemother. 2017 Jul 25;61(8).

36. Gabrielaite M, Johansen HK, Molin S, Nielsen FC, Marvig RL. Gene loss and acquisition in lineages of bacteria evolving in a human host environment. BioRxiv. 2020 Feb 3;

37. Marvig RL, Damkiær S, Khademi SM, Markussen TM, Molin S, Jelsbak L. Within-host evolution of *Pseudomonas aeruginosa* reveals adaptation toward iron acquisition from hemoglobin. MBio. 2014 May 6;5(3):e00966-14.

38. Tyrrell J, Whelan N, Wright C, Sá-Correia I, McClean S, Thomas M, et al. Investigation of the multifaceted iron acquisition strategies of Burkholderia cenocepacia. Biometals. 2015 Apr;28(2):367–80.

39. Hu Y, Zhu Y, Ma Y, Liu F, Lu N, Yang X, et al. Genomic insights into intrinsic and acquired drug resistance mechanisms in Achromobacter xylosoxidans. Antimicrob Agents Chemother. 2015 Feb;59(2):1152–61.

40. Sommer LM, Johansen HK, Molin S. Antibiotic resistance in Pseudomonas aeruginosa and adaptation to complex dynamic environments. Microb Genom. 2020 Apr 29;

41. Morris RP, Nguyen L, Gatfield J, Visconti K, Nguyen K, Schnappinger D, et al. Ancestral antibiotic resistance in Mycobacterium tuberculosis. Proc Natl Acad Sci USA. 2005 Aug 23;102(34):12200–5.

42. Lieberman TD, Flett KB, Yelin I, Martin TR, McAdam AJ, Priebe GP, et al. Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. Nat Genet. 2014 Jan;46(1):82–7.

43. Lieberman TD, Wilson D, Misra R, Xiong LL, Moodley P, Cohen T, et al. Genomic

diversity in autopsy samples reveals within-host dissemination of HIV-associated Mycobacterium tuberculosis. Nat Med. 2016 Oct 31;22(12):1470–4.

44. Gabrielaite M, Marvig RL. Pactyper: Snakemake pipeline for continuous clone type prediction for WGS sequenced bacterial isolates based on their core genome. Zenodo. 2020;

45. Gabrielaite M, Marvig RL. GenAPI: a tool for gene absence-presence identification in fragmented bacterial genome sequences. BioRxiv. 2019 Jun 3;

46. Gabrielaite M, Marvig RL. GenAPI: a tool for gene absence-presence identification in fragmented bacterial genome sequences. BMC Bioinformatics. 2020 Dec;21(1):320.

47. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012 May;19(5):455–77.

48. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 2014 Jan 1;15(11):524.

49. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019 Jul 2;47(W1):W256–9.

50. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016 Jan 4;44(D1):D733-45.

51. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014 May 1;30(9):1312–3.

52. Seemann T. Snippy: Rapid haploid variant calling and core genome alignment [Internet]. 2018 [cited 2020 Feb 10]. Available from: https://github.com/tseemann/snippy

53. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom. 2016 Nov 30;2(11):e000093.

54. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018 Nov 30;9(1):5114.

55. R: The R Project for Statistical Computing [Internet]. [cited 2020 Feb 10]. Available from: https://www.r-project.org/

56. CRAN - Package pheatmap [Internet]. [cited 2020 Feb 10]. Available from: https://cran.r-project.org/web/packages/pheatmap/index.html

57. Gabrielaite M, Misiakou M-A, Marvig RL. BacDist: Snakemake pipeline for bacterial SNP distance and phylogeny analysis. Zenodo. 2020;

58. Denamur E, Matic I. Evolution of mutation rates in bacteria. Mol Microbiol. 2006

May;60(4):820–7.

59. Barrick JE, Lenski RE. Genome dynamics during experimental evolution. Nat Rev Genet. 2013 Dec;14(12):827–39.

60. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2014 Apr 10;10(4):e1003537.

61. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using tracer 1.7. Syst Biol. 2018 Sep 1;67(5):901–4.

62. Seemann T. Abricate: Mass screening of contigs for antimicrobial and virulence genes [Internet]. 2018 [cited 2020 Feb 10]. Available from: https://github.com/tseemann/abricate

63. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res. 2016 Jan 4;44(D1):D694-7.

64. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012 Nov;67(11):2640–4.

65. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol. 2017 Aug 1;34(8):2115–22.

66. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012 Dec 1;28(23):3150–2.

67. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15;30(14):2068–9.

68. Jaillard M, Lima L, Tournoud M, Mahé P, van Belkum A, Lacroix V, et al. A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events. PLoS Genet. 2018 Nov 12;14(11):e1007758.

69. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019 Jan 8;47(D1):D506–15.

70. Microreact - Phylogenetic tree of clinical Achromobacter isolates and Achromobacter genomes available on RefSeq database [Internet]. [cited 2020 Jun 23]. Available from: https://microreact.org/project/XBjPPEXWl

71. Microreact - 101 Achromobacter clinical isolate phylogenetic tree [Internet]. [cited 2020 Jun 23]. Available from: https://microreact.org/project/ByZx4dqC7

72. iTOL - Phylogenetic tree of DES isolates [Internet]. [cited 2020 Jun 23]. Available from: https://itol.embl.de/tree/1280736237908158497736 2