ACCESS MICROBIOLOGY

RESEARCH ARTICLE
Fischer *et al.*, *Access Microbiology* 2021;3:000286
DOI 10.1099/acmi.0.000286

MICROBIOLOGY SOCIETY

OPEN ACCESS

Sequence diversity of the *Pseudomonas aeruginosa* population in loci that undergo microevolution in cystic fibrosis airways

Sebastian Fischer[1]†, Jens Klockgether[1]†, Marina Gonzalez Sorribes[1,2], Marie Dorda[1,3], Lutz Wiehlmann[1,3] and Burkhard Tümmler[1,4,*]

Abstract

Five hundred and thirty-four unrelated *Pseudomonas aeruginosa* isolates from inanimate habitats, patients with cystic fibrosis (CF) and other human infections were sequenced in 19 genes that had been identified previously as the hot spots of genomic within-host evolution in serial isolates from 12 CF lungs. Amplicon sequencing confirmed a significantly higher sequence diversity of the 19 loci in *P. aeruginosa* isolates from CF patients compared to those from other habitats, but this overrepresentation was mainly due to the larger share of synonymous substitutions. Correspondingly, non-synonymous substitutions were either rare (*gltT*, *lepA*, *ptsP*) or benign (*nuoL*, *fleR*, *pelF*) in some loci. Other loci, however, showed an accumulation of non-neutral coding variants. Strains from the CF habitat were often mutated at evolutionarily conserved positions in the elements of stringent response (RelA, SpoT), LPS (PagL), polyamine transport (SpuE, SpuF) and alginate biosynthesis (AlgG, AlgU). The strongest skew towards the CF lung habitat was seen for amino acid sequence variants in AlgG that clustered in the carbohydrate-binding/sugar hydrolysis domain. The master regulators of quorum sensing *lasR* and *rhlR* were frequent targets for coding variants in isolates from chronic and acute human infections. Unique variants in *lasR* showed strong evidence of positive selection indicated by d_N/d_S values of ~4. The *pelA* gene that encodes a multidomain enzyme involved in both the formation and dispersion of Pel biofilms carried the highest number of single-nucleotide variants among the 19 genes and was the only gene with a higher frequency of missense mutations in *P. aeruginosa* strains from non-CF habitats than in isolates from CF airways. PelA protein variants are widely distributed in the *P. aeruginosa* population. In conclusion, coding variants in a subset of the examined loci are indeed characteristic for the adaptation of *P. aeruginosa* to the CF airways, but for other loci the elevated mutation rate is more indicative of infections in human habitats (*lasR, rhlR*) or global diversifying selection (*pelA*).

DATA SUMMARY

The Illumina (NGS) sequence dataset generated and used in this study is available in the European Nucleotide Archive (ENA) with study accession number PRJEB45250, hosted by the European Bioinformatics Institute EMBL-EBI (www.ebi.ac.uk). Relevant code was made available through https://githubcom/kfgpfz/workflow_amplicons_sequence_diversity. The authors confirm that all supporting data, code, protocols and accession numbers have been provided within the article and through supplementary data files. The genome sequencing datasets of the 262 serial *P. aeruginosa* isolates from the 12 people with cystic fibrosis are deposited in the ENA with study accession no. PRJEB5303. The genome sequences of *P. aeruginosa* PA14 and 19 other strains from the environment (Table S2, available in the online version of this article) were retrieved from the National Center for Biotechnology Information (NCBI).

INTRODUCTION

Cystic fibrosis (CF) is a life-limiting autosomal recessive trait that is caused by mutations in the *cystic fibrosis transmembrane*

Received 17 June 2021; Accepted 04 October 2021; Published 07 December 2021
Author affiliations: [1]Clinical Research Group 'Molecular Pathology of Cystic Fibrosis', Department for Paediatric Pneumology, Allergology and Neonatology, Hannover Medical School, Hannover, Germany; [2]Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain; [3]Research Core Unit Genomics, Hannover Medical School, Hannover, Germany; [4]Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH), German Center for Lung Research, Hannover Medical School, Germany.
*Correspondence: Burkhard Tümmler, tuemmler.burkhard@mh-hannover.de
Keywords: amplicon sequencing; cystic fibrosis; niche adaptation; population genetics; *Pseudomonas aeruginosa*.
Abbreviations: AHL, *N*-acyl homoserine lactone; CF, cystic fibrosis; CFTR, Cystic Fibrosis Transmembrane Conductance Regulator; C_4-HSL, *N*-butanoyl-homoserine lactone; COPD, chronic obstructive pulmonary disease; ENA, European Nucleotide Archive; NGS, next generation sequencing; 3-oxo-C_{12}-HSL, *N*-3-oxo-dodecanoyl-homoserine lactone; PCR, polymerase chain reaction; SNV, single nucleotide variant; SRA, Sequence Read Archive.
†These authors contributed equally to this work
Supplementary materials are available with the online version of this article.

conductance regulator (CFTR) gene [1]. The basic defect of impaired luminal secretion of chloride and bicarbonate [2] predisposes individuals to chronic airway infections with opportunistic pathogens, namely *Staphylococcus aureus* and later in life *Pseudomonas aeruginosa,* which determine the quality of life and prognosis for most persons with CF [3–5].

Once a *P. aeruginosa* clone has colonized its niche in the upper and lower CF airways, it will diversify in genotype and phenotype and will persist for decades unless it is replaced by another more proficient *P. aeruginosa* clone [6, 7]. Antimicrobial chemotherapy is successful in eradicating *P. aeruginosa* from CF patients' airways during the initial phase of colonization; later on, however, systemic or topical application of antipseudomonal agents may suppress the bacterial load, but will not eliminate the *P. aeruginosa* clone [8, 9].

Chronic, or even life-long, CF airway infections with *P. aeruginosa* provide a rare opportunity to study long-term bacterial microevolution in a human host [10, 11]. To understand how the aquatic environmental micro-organism *P. aeruginosa* conquers the CF airways, adapts its lifestyle to an atypical habitat and escapes the CF host's immune response, the genomic adaptation of *P. aeruginosa* to the CF lungs has been investigated by whole-genome sequencing of serial isolates from patients' lungs. The genomic evolution of *P. aeruginosa* has initially been studied in serial isolates from the same patient [12–16] or in transmissible lineages between different patients [17, 18], but more recently has been analysed in numerous patients seen at the same CF clinic [19–21].

We followed the microevolution of the first persisting *P. aeruginosa* clone by whole-genome sequencing of serial isolates from highly divergent clinical courses of airway infection at the CF clinic Hannover, Germany, i.e. six courses from the most severely affected CF patients with a fatal outcome because of respiratory insufficiency within less than 15 years, and six courses from mildly affected patients with a rather normal daily life 25–35 years after acquisition of *P. aeruginosa* [21]. Thereby we identified gene loci that were repetitively hit in the 12 different *P. aeruginosa* clones. When we sorted these frequently affected genes by the severity of mutation, loss-of-function mutations and drastic missense mutations were predominantly found in transporters, the alginate biosynthesis operon, sensors of environmental cues, regulators of quorum sensing and key players of the stringent response. These gene loci should qualify as the hot spots of pathoadaptive mutations in CF lungs. However, this interpretation resides on a dataset generated from just 12 CF patients' isolates. Only a few loci such as *lasR* or *algU* had already been identified in CF isolates of diverse geographical origin as hot spots of mutation [22–26]. Thus the top candidates in our genomic microevolution study were selected to be validated in a comprehensive strain collection to determine whether they represent prime targets of mutation of *P. aeruginosa* thriving in the CF lung habitat. Alternatively, these loci could also be frequently mutated in any acute or chronic human infection, with CF just being an example, or they could present a prime target of diversifying selection in any animate and inanimate habitat. To differentiate between these various alternatives, we sequenced the putative hot spots of mutation in a representative strain collection of 345 unrelated CF isolates from diverse CF clinics, 92 isolates from other acute and chronic human infections and 97 isolates from the inanimate environment [27, 28]. Sequence diversity turned out to be significantly higher among the strains from CF lungs, but this general finding was attributed to the larger portion of silent mutations. The frequency and severity of missense mutations, however, varied by locus and was a characteristic feature of the CF habitat for only 5 of the 19 tested loci. Genes that are mutated frequently during the colonization of CF lungs do not implicitly qualify to represent the CF habitat-specific genomic adaptation of *P. aeruginosa*.

Impact Statement

A study of the within-host evolution of *P. aeruginosa* in 12 people with CF identified gene loci that were repetitively hit by mutations. In this work, these loci were sequenced in hundreds of *P. aeruginosa* isolates from (a) unrelated CF patients seen at diverse clinics, (b) other human infections and (c) the environment in order to clarify whether or not mutations in these loci are characteristic for the microevolution of *P. aeruginosa* in CF lungs. Global sequence diversity was indeed higher in CF than in non-CF isolates for all examined loci, but the vast majority of the nucleotide substitutions do not change the coding sequence. CF-associated coding variants mainly affected loci that shape the phenotypic conversion of *P. aeruginosa* in CF lungs, i.e. LPS deficiency, biofilm formation and alginate production. The 'hot spot of non-synonymous mutation' turned out to be AlgG, which converts mannuronic acid into α-L-guluronate during alginate biosynthesis. Missense or nonsense mutations in the master regulators of quorum sensing are common in *P. aeruginosa* isolates from CF lungs, but this comparative study clarified that *lasR* and *rhlR* mutations frequently emerge in all kinds of human infection.

METHODS

P. aeruginosa strains

The *P. aeruginosa* isolates used in his study were taken from various strain collections, of which the majority had already been genotyped within an earlier population study [27]. If not, the clonal lineage of isolates was analysed with the same binary marker microarray as before, resulting in a four-digit code describing the respective clonal lineage of an isolate [29]. In total, 534 isolates were selected: 345 from a CF strain collection, 25 from a COPD strain collection, 67 from acute infection scenarios and 97 of environmental origin (see Table S1). If more than one serial isolate from a CF or COPD patient was available, one representative of each detected clonal lineage was included in the study, reducing redundancy by excluding non-independent serial isolates. Hypermutable strains were excluded. Strains were stored at −80 °C in culture medium [Luria–Bertani broth (LB)] supplemented with 15% (v/v) glycerol. Strains were cultivated in 5 ml liquid LB medium at

37 °C under constant shaking for 16 h for subsequent DNA isolation.

Genomic DNA isolation

Bacterial cells were harvested from 1 to 2 ml liquid culture, and the genomic DNA was extracted and purified either by cell lysis and phenol–chloroform–isoamyl alcohol extraction [30], or by using the Qiagen genomic DNA extraction kit for microbial cells. DNA was quantified and checked for purity and integrity by spectrophotometry at 260 and 280 nm, and by Qubit fluorometry using the ds DNA HS assay kit (Thermo Fisher Scientific).

Primer design and PCR

In this study, 19 genomic loci were analysed by PCR amplification and subsequent sequencing. As blueprints, the sequences of the respective genes were taken from the genome of the *P. aeruginosa* reference strain PA14 (GenBank accession no. NC_008463.1). Depending on the gene size, primer pairs for one to three overlapping PCR products were designed that cover the genes' coding sequences completely. Sequences were selected with the help of the tool Primer 3 (http://primer3.ut.ee) in order to guarantee highly similar hybridization temperatures for all primers. Primer sequences were tested and modified, if necessary, for stable generation of the expected products. In total, 31 primer pairs for the 19 loci were chosen that would generate PCR products with sizes between 582 and 1528 bp (see Table S3). For multiplex PCRs, equal amounts of genomic DNA from 18 to 25 isolates each were combined in so-called 'template pools' (see Table S1) to amplify a target sequence for several strains within a single PCR reaction. Thus, 17 template pools were generated with DNA of CF isolates and 5, 1 and 3 pools with DNA from environmental, COPD and acute infection isolates, respectively. Aliquots from each DNA pool were then used as templates for PCR reactions with each primer pair.

A total of 806 PCR reactions were performed in 96-well plates using Goldstar *Taq* polymerase (Eurogentec) and the supplied reaction buffer. Ten nanograms of template DNA and 5% (v/v) DMSO were added to the reaction mix, as well as 0.67 µl 25 mM $MgCl_2$ solution, 2 µl each of 5 µM stock solutions of the respective primers and water dH_2O to a final volume of 20 µl per reaction. The PCR protocol comprised 30 amplification cycles with time periods of 60 s at 60 °C for primer annealing, 60 s at 72 °C for elongation and 90 s at 94 °C for DNA denaturation. The yield and purity of the PCR products were checked subsequently by agarose gel electrophoresis. PCR reactions with insufficient results were repeated in single tubes with a modified protocol. Reaction mixes with overall volumes of 25 µl included 0.8 µl 25 mM $MgCl_2$ solution and 6.25 µl of a 4 M solution of the PCR enhancer betaine instead of DMSO. These reactions were performed at 58 or 59 °C as annealing and 95 °C as denaturation temperature, and the initial denaturation was performed at 96 °C for 300 s. After generating sufficient amounts of PCR product for all template–primer combinations, aliquots of all 31 reactions from the same template pool were combined, generating 26 amplicon pools for DNA library preparation and sequencing. The respective volumes of the single PCR aliquots were adjusted upon visual inspection of band intensities in the agarose gel controls in order to provide comparable amounts of each product in the mixture taken for the sequencing.

DNA library preparation and sequencing

The DNA concentration of each amplicon pool was measured at the Qubit and adjusted to 1 ng µl^{-1}. An aliquot of 130 µl was sheared in a Covaris S2 system to generate DNA fragments of ~300 bp in length. For the generation of fragment libraries, 50 µl of the sheared DNA solution was purified with 1.5 vol. Agencourt AMPureXP (A63881, Beckman Coulter). Twenty nanograms of purified DNA were used to prepare fragment libraries with the NEBNext Ultra II DNA Library Prep kit for Illumina (E7645L, New England Biolabs) and NEBNext unique dual index primer pairs (E6440L, New England Biolabs). For amplification, eight PCR cycles were applied. DNA fragment sizes of fragment libraries were monitored using the Bioanalyzer High Sensitivity DNA assay (5067–4626, Agilent Technologies). Equimolar amounts of individually barcoded libraries were pooled. The library pool was denatured with NaOH and finally diluted to 10 pM. Sequencing was performed on an Illumina MiSeq system (MiSeq Reagent kit v2 Micro; 2×151 bp paired end reads, 2×8 bp indices and 1% control v3 PhiX) and generated between 113687 and 269991 read pairs for each of the 26 amplicon pools.

Sequence analysis

These datasets were initially processed by removing adapter sequences and clearing low-quality reads with the tool Trimmomatic [31]. The processed data were then aligned as single-end reads to the *P. aeruginosa* PA14 reference sequence using bwa mem [32]. For each genome position of the 19 loci and each sequencing pool the coverage and proportion of variants were calculated for each position. The cutoff criteria for calling a single-nucleotide variant (SNV) candidate were a minimum coverage of 200, a variant call of at least 20 reads and a proportion of 2.5% of all reads. The number of isolates per pool encoding a variant was calculated for all candidate positions by the integer of the ratio of the number of variant reads to the total number of reads times the number of isolates. Next, the results for each pool were merged by habitat using an in-house R script. SNP statistics such as synonymous and non-synonymous exchanges were processed using SNPeff [33].

Validation of SNVs

SNV predictions derived from the batch amplicon sequencing data were validated by diagnostic restriction digests. In parallel, PCR products of single isolates were digested in order to determine the frequency of an SNV in the respective isolate pool. For these analyses 15 SNVs in *algG*, *algU*, *lasR* or *pagL* were selected, which were predicted to be present in one to three isolate pools. Positions were chosen that were covered by a high (>1000) or low (<200) numbers or reads and/or by an equal (<1.5:1) or unequal (>2:1) number of reads on both strands. For the digests, DNA flanking the potential SNV position was separately amplified from genomic DNA of each strain of a pool. If informative restriction enzymes differentiating wild-type from SNV DNA could be identified, the amplicons of the respective locus and of the whole batch were digested in parallel. Otherwise, restriction recognition sites, including the SNV position, were artificially generated by introducing nucleotide changes in close vicinity with one of the PCR primers. In such cases, PCR products of 104–128 bp were generated with either GoldStar *Taq* polymerase or Q5 High Fidelity DNA polymerase (New England Biolabs). Restriction digests were performed with 0.2–1 µg PCR product per reaction using enzymes and the corresponding reaction buffers according to the manufacturer's instructions (New England Biolabs). The digests were checked on agarose gels for fragment patterns indicating the wild-type sequence or the SNV containing sequence for each strain.

RESULTS AND DISCUSSION

Selection of the gene loci

We have collected serial *P. aeruginosa* isolates at half-yearly intervals until now from the respiratory secretions of 32 individuals with CF seen regularly at the CF clinic Hannover, who had become chronically colonized with *P. aeruginosa* between 1984 and 1992 [34]. Whole *P. aeruginosa* genome sequencing was performed on serial isolates from the six patients with the most severe clinical courses and the six patients with the mildest courses of chronic lung infection [21]. We examined the microevolution of the first persisting clone from its acquisition until the patient's death or its permanent replacement by another clone. The annual rate of bacterial mutations per megabase varied between 1.2 and 1.8 in the absence and between 6.2 and 80.7 in the presence of hypermutable strains [21]. Of the hundreds of gene loci that were hit by frameshift mutation or non-synonymous nucleotide substitutions, we now selected for this study the most frequently mutated loci for sequence analysis of *P. aeruginosa* isolates from the inanimate environment, CF patients seen at other CF centres, patients with COPD and acute infections in humans and livestock (Table S1). A gene locus of the *P. aeruginosa* core genome was assigned to the group of most frequently mutated loci during microevolution in CF lungs if mutations had affected the locus in at least 3 of the 12 lungs, had persisted in a clade and had also occurred in non-hypermutable lineages. Table 1 lists the 19 protein-coding loci that fulfilled these criteria. Now we explored in this study whether these loci were only hot spots of mutation in the Hannover cohort of CF patients or whether they were frequently mutated in the global CF community and/or other common habitats of the *P. aeruginosa* population.

Strain panel and sequencing

The strain panel consisted of 345 airway isolates from patients with CF, 25 airway isolates from patients with COPD, 67 isolates from acute infections and 97 isolates from inanimate predominantly aquatic habitats (Table S1). Batches of 18 to 25 isolates, each belonging to the same habitat, were subjected to deep multiplex amplicon sequencing of the 19 target genes. Amplicon profiles of coverage were matching among the 26 batches but the coverage of individual gene segments varied substantively from 200- to 10 000-fold. The expected median coverage at each nucleotide position had *a priori* been set to 1000-fold, the minimal coverage was set to 200-fold. Gene segments were resequenced if the coverage was below this threshold.

These guidelines were the spin-off of the experimental validation of nucleotide substitutions in the primary sequence data by informative restriction digestion analysis. Selected SNVs predicted from the amplicon sequencing results were queried by restriction digests by testing for SNV-bearing recognition sites within PCR products generated from the genomic DNAs of the respective batch of isolates (see Methods). The results for 20 candidate SNV predictions confirmed the presence of a sequence variation with at least 90% probability after we had excluded all candidate positions with read coverages below 200 and fewer than 20 SNV reads and had excluded all positions with unusual patterns in the sequence read alignment that indicated alignment or amplification errors as potential causes for false-positive SNV predictions.

Sequence diversity among the 19 loci

Amplicon sequencing of the 19 genes in the strain panel identified SNVs at 6.0–14.4% of nucleotide positions compared to the strain PA14 reference. The average normalized frequency of SNVs was 8.4% (Table 2). If we only counted the nucleotide substitutions in the strain panel that were absent in the genomes of 19 isolates from environmental habitats (Table S2), the frequency of sequence variants decreased in all genes (average 6.5%, range 4.1–12.6%). SNVs that were only detected in 1 of the 534 strains, occurred with an average frequency of 2.1% (range 1.2–7.8%) (Table 2). Sequence variants were significantly more abundant in the CF isolates than in the whole strain panel (Table 3). This statement applies to all 19 tested loci. Thus, these 19 genes that were identified as hot spots of mutation in serial airway isolates from 12 CF patients seen at a single clinic are indeed loci that mutate more frequently in CF lungs than in other habitats of the *P. aeruginosa* population. Sequence diversity was also significantly higher than average among the environmental isolates for 7 of the 19 genes, namely *algG*, *gacA*, *lepA*, *nuoL*, PA4391, *pelA*, *spuF* (Table 3).

Next, we calculated d_N/d_S of the 19 target loci in CF and non-CF *P. aeruginosa* isolates, i.e. the ratio of non-synonymous substitutions per non-synonymous sites to the number of synonymous substitutions per synonymous sites standardized to the codon usage in *P. aeruginosa* (Table 4).

Table 1. *P. aeruginosa* genes that were identified as hot spots of mutation in serial isolates from cystic fibrosis patients seen at the CF clinic Hannover, Germany.

Gene			
Name	**Locus**	**Length (bp)**	**Annotation**
algG	PA3545	1632	Alginate-c5-mannuronan-epimerase AlgG
algU	PA0762	582	Sigma factor AlgU
fleR	PA1099	1422	Two-component response regulator (of motility and adhesion to mucin)
gacA	PA2586	645	Response regulator GacA
gltR	PA3192	729	Two-component response regulator GltR (to presence of glucose)
lasR	PA1430	720	Transcriptional regulator LasR
lepA	PA0767	1800	GTP-binding protein LepA
nuoL	PA2647	1848	NADH dehydrogenase I chain L
	PA4391	1002	Hypothetical protein (orthologue in PA14 strain: PA14_57070)
	PA5048	768	Probable nuclease (orthologue in PA14 strain: PA14_66700)
pagL	PA4661	522	Lipid A 3-O-deacylase
pelA	PA3064	2847	PelA (multi-domain enzyme with PEL deacetylase and hydrolase activities)
pelF	PA3059	1524	PelF (UDP-GalNAc/GlcNAc-glycosyltransferase)
ptsP	PA0337	2280	Phosphoenolpyruvate-protein phosphotransferase PtsP
relA	PA0934	2244	GTP pyrophosphokinase
rhlR	PA3477	726	Transcriptional regulator RhlR
spoT	PA5338	2106	Guanosine-3′,5′-bis(diphosphate) 3′-pyrophosphohydrolase
spuE	PA0301	1098	Polyamine transport protein 8 (spermidine-binding protein)
spuF	PA0302	1155	Polyamine transport protein PotG (ABC transporter ATPase)

Values of d_N/d_S of ~1 indicate neutral evolution, values >1 positive selection and values <1 purifying selection towards synonymous nucleotide substitutions. Whole-genome comparisons of major clones in the *P. aeruginosa* population revealed an average d_N/d_S value of 0.1 [21].

The panel of sequence variants in both CF and non-CF isolates showed d_N/d_S values far below 1, implying that sequence diversity is dominated by synonymous substitutions in all 19 loci (Table 4). The individual d_N/d_S ratios were, moreover, below 0.1 in all but two loci, indicating that the hot spots of mutation in CF lungs are under stronger purifying selection than the average gene locus in the global *P. aeruginosa* population. When we removed all SNVs from the list that were also present in the selected set of reference genomes from the environment (Table S2), d_N/d_S increased in all but two loci on average by approximately fivefold to a maximum value of 0.68 (Table 4), implying that amino acid sequence variants are more frequent among the less common SNVs. The loci associated with alarmone (*spoT*), LPS (*pagL*) or alginate biosynthesis (*algG*, *algU*) and the master regulator *gacA* exhibited at least twofold higher d_N/d_S values in the collection of isolates from CF lungs than in those from other habitats. These loci are known to be key for *P. aeruginosa* to establish its lifestyle in CF airways [11].

The very same loci plus *fleR, relA* and *spuE* also exhibited at least twofold higher d_N/d_S ratios in CF than in non-CF among the unique SNVs that had been identified in only a single isolate of the panel of 534 strains (Table 4). Compared to all SNVs in the 19 genes, the d_N/d_S ratios were on average 18.2-fold and 9.3-fold higher among the unique SNVs in CF and non-CF strains, respectively, demonstrating a substantively larger share of nonsynonymous substitutions among the rare or *de novo* SNVs. The larger fraction of amino acid sequence variants in CF compared to non-CF isolates supports the conclusion drawn from the genome sequencing of serial CF isolates that the hot spots of mutation play a prominent role in the adaptation of *P. aeruginosa* to colonize and persist in CF airways [21]. A d_N/d_S ratio above 1 that indicates Darwinian positive selection was calculated for the unique SNVs of the CF strain panel in the sigma factor *algU* and the transcriptional regulators *rhlR* and *lasR. AlgU* is a target for the conversion to non-mucoidy by second site suppressor mutations [22, 35], and *rhlR* and *lasR* code for key elements of the quorum sensing signalling network [36, 37]. LasR variants are frequent in the

Table 2. Normalized frequency of sequence variants in the strain panel (%)*

Gene	Compared to PA14 reference genome	Absent in a reference panel of environmental strains	Unique in one strain
algG	8.4	6.4	2.2
algU	10.8	8.9	4.1
fleR	10.5	7.6	2.7
gacA	7.1	5.7	3.4
gltR	8.9	6.0	1.1
lasR	14.4	12.6	7.5
lepA	7.3	4.7	1.8
nuoL	6.6	4.1	1.4
PA4391	9.6	8.0	2.1
PA5048	12.1	9.0	2.2
pagL	9.2	7.5	3.8
pelA	10.4	7.5	2.3
pelF	11.7	9.3	1.9
ptsP	6.4	4.2	1.2
relA	9.4	6.6	1.6
rhlR	8.3	5.9	3.0
spoT	6.0	4.1	1.7
spuE	8.4	6.1	2.2
spuF	6.0	4.9	1.4

*The frequency of sequence variants as a percentage was determined by (the total number of sequence variants)/(gene length×number of isolates), i.e. hits at the same position were added by counts.

Table 3. Habitat-associated abundance of sequence variants in *P. aeruginosa* isolates*

Gene	Environment	Acute infection	COPD	CF
algG	+	−	−	+
algU	−	−	−	+
fleR	−	−	−	+
gacA	+	−	−	+
gltR	−	−	−	+
lasR	−	−	−	+
lepA	+	−	−	+
nuoL	+	−	−	+
PA4391	+	−	−	+
PA5048	−	−	−	+
pagL	−	−	−	+
pelA	+	−	−	+
pelF	−	−	−	+
ptsP	−	−	−	+
relA	−	−	−	+
rhlR	−	−	−	+
spoT	−	−	−	+
spuE	−	−	−	+
spuF	+	−	−	+

+, higher abundance than average; −, lower abundance than average.

*The normalized frequency distributions of sequence variants in each selected *P. aeruginosa* locus were compared between isolates from the environment, acute infection and airways of patients with COPD or CF. The frequency distribution of sequence variants was significantly different in the four habitats for all tested genes (Bonferroni-corrected $P_{corr} < 0.05$). Only sequence variants were considered that are absent in the reference panel of environmental isolates with completely sequenced genomes (see Table S2).

CF habitat [12, 25] and give rise to diverse phenotypes [25]. The d_N/d_S ratio of 4.2 for the natural CF isolates is higher than the highest values reported in the literature for a *P. aeruginosa* gene to date, i.e. *fpvA* and *fpvG* involved in ferripyoverdine transport [38, 39]. In other words, *de novo* non-synonymous mutations in *lasR* are strongly selected in the niche of the CF lungs. *lasR* was, moreover, the only gene in our panel among the non-CF isolates that showed positive selection, with a remarkable d_N/d_S ratio of 2.7. In summary, the emergence of non-synonymous mutations in *lasR* is common in the global *P. aeruginosa* population.

The frequency distribution of amino acid sequence variants was comparable in CF and non-CF strain subpopulations for 14 of the 19 genes (Table 5). AlgG, AlgU, PagL and SpoT missense variants, however, were significantly more common in isolates from CF airways, whereas PelA variants were more common in non-CF isolates (Table 5). As indicated by the higher Dayhoff score [40], the non-synonymous substitutions shared by the CF and non-CF strain populations were more benign than the amino acid changes observed solely in CF or non-CF isolates ($P<10^{-12}$). CF and non-CF isolates did not differ in their spectrum of amino acid substitutions (Table 6, Fig. 1).

Sequence diversity of individual loci

After having provided some global features of the sequence diversity among the 19 loci in our strain panel, we now describe aspects of the mutation spectrum in individual loci. The majority of targets are transcriptional regulators that respond to signals and environmental cues with the activation of transcriptional programmes (Fig. 2). Alternatively, translation is modulated (*relA, spoT*). The second largest

Table 4. Ratio of nonsynonymous to synonymous substitutions d_N/d_S in the examined genes*

Gene	**Substitutions compared to PA 14 reference sequence**		**Substitutions absent in the genomes of environmental strains (Table S2)**		**Unique substitutions in one strain**	
panels	**CF**	**non-CF**	**CF**	**non-CF**	**CF**	**non-CF**
algG	0.14	0.11	0.83	0.44	0.87	0.075
algU	0.0085	0.0005	0.054	0.005	1.73	0.081
fleR	0.03	0.02	0.11	0.08	0.40	0.13
gacA	0.004	0.002	0.046	0.020	0.91	0.055
gltR	0.008	0.006	0.082	0.097	–/–	0.36
lasR	0.023	0.017	0.22	0.35	4.24	3.91
lepA	0.007	0.007	0.043	0.034	0.089	0.12
nuoL	0.085	0.090	0.11	0.11	0.21	0.19
PA4391	0.049	0.038	0.32	0.24	0.16	0.33
PA5048	0.084	0.072	0.12	0.24	0.15	0.25
pagL	0.044	0.035	0.40	0.15	1.51	0.12
pelA	0.135	0.139	0.23	0.27	0.52	0.36
pelF	0.084	0.072	0.16	0.18	0.25	0.24
ptsP	0.019	0.054	0.021	0.055	0.11	–/–
relA	0.006	0.003	0.028	0.026	0.34	0.020
rhlR	0.008	0.007	0.056	0.076	1.44	0.86
spoT	0.004	0.001	0.061	0.022	0.32	0.029
spuE	0.025	0.033	0.099	0.14	0.44	0.11
spuF	0.019	0.006	0.056	0.034	0.11	–/–

*d_N/d_S, i.e. the ratio of nonsynonymous substitutions per nonsynonymous sites to the number of synonymous substitutions per synonymous sites. Whole-genome comparisons of unrelated clonal *P. aeruginosa* complexes revealed an empirical median d_N/d_S value of 0.1 [21, 62].

group of targets are cell wall constituents that are involved in the uptake of polyamines (*spuE, spuF*) or the production of exopolysaccharides (*algG*, *pelA*, *pelF*). Singular targets contribute to energy production (*nuoL*), coordination of metabolism (*ptsP*), ribosome biogenesis (*lepA*), or modify lipopolysaccharide structure (*pagL*) (Fig. 2). We now focus on the 11 loci with high sequence diversity, positive selection of unique SNVs or preponderance of missense mutations in isolates from the CF airway habitat. The mutation spectrum of six other loci is discussed in the Document S1.

RelA and *SpoT*

The stringent response proteins RelA and SpoT control the cellular levels of the second messenger alarmone (p)ppGpp [41]. RelA binds uncharged tRNAs, the tRNA–RelA complex is then loaded into the ribosomal A-site and the interaction with the 23S rDNA activates RelA to synthesize (p)ppGpp [42, 43]. Whereas RelA lacks hydrolytic activity, the bifunctional enzyme SpoT is a weak synthetase and strong hydrolase of ppGpp [44]. The SpoT amino acid sequence is conserved in the global *P. aeruginosa* population. Only three benign polymorphisms at non-conserved positions were observed among the non-CF isolates of our strain panel. In contrast, 17 coding variants were confined to the CF isolates of which 12 substitutions were detected in single isolates, 11 substitutions affected positions that are conserved among SpoT orthologs and 9 non-conservative substitutions changed the polarity or charge of the amino acid residue (Table S4). SpoT mutants were found in 13% of the strains. RelA has a two-domain architecture: an N-terminal domain bearing hydrolase and synthetase subdomains and a C-terminal domain bearing three subdomains that are critical for ribosome binding and autoinhibition of synthetase activity in the absence of the ribosome. We found 190 synonymous and 22 non-synonymous substitutions in the *relA* gene of our strain panel that were strongly overrepresented in the CF isolates ($P_{\mathrm{corr}}=7\times10^{-45}$) (Table S4). Non-conservative mutations probably relevant for RelA function were observed at 12 positions, of which 6 were located in 1 of the subdomains. Thirty-seven strains (6.9%) were carrying a stop mutation in *relA*, indicating that RelA knock-out mutants constitute a few per cent of the natural *P. aeruginosa* population. In summary, SpoT and/or

Table 5. Frequency of amino acid sequence variants in the *P. aeruginosa* strain panel

Gene	No. of missense mutations in CF and non-CF isolates			Ratio	Fisher's test
	Common in CF and non-CF	CF only	Non-CF only	CF/non-CF*	Presence of CF vs non-CF
algG	20	47	4	6.3	*P*=0.006
algU	0	23	1	12.3	*P*=0.01
fleR	25	21	4	2.8	*P*=0.07
gacA	0	12	2	3.2	*P*=0.15
gltR	2	4	4	0.50	*P*=0.34
lasR	1	47	28	0.90	*P*=0.13
lepA	6	3	7	0.23	*P*=0.15
nuoL	15	10	7	0.77	*P*=0.26
PA4391	17	10	13	0.41	*P*=0.08
PA5048	16	8	5	0.85	*P*=0.31
pagL	3	19	1	10.1	*P*=0.02
pelA	63	25	29	0.46	*P*=0.02
pelF	42	10	10	0.53	*P*=0.16
ptsP	4	6	0		*P*=0.23
relA	3	16	3	2.8	*P*=0.10
rhlR	0	14	9	0.83	*P*=0.23
spoT	2	17	1	9.1	*P*=0.04
spuE	10	7	4	0.93	*P*=0.34
spuF	3	5	0		*P*=0.22

*Normalized ratio of variants detected in 345 CF and 184 non-CF *P. aeruginosa* isolates.

RelA variants were found in approximately one-third of CF isolates. At least the drastic amino acid substitutions should modulate the diverse roles of the two proteins in stringent response, normal growth and virulence of *P. aeruginosa* [44].

SpuE and *SpuF*

The *spuDEFGH* gene cluster encodes an ABC transporter for polyamines. The *spuE* ($P_{corr}=4\times10^{-8}$) and *spuF* ($P_{corr}=0.01$) genes were confirmed to be hot spots of mutation in CF airway isolates of *P. aeruginosa* (Table S4). The majority of the amino acid substitutions in SpuE changed polarity or charge, affected positions conserved in bacterial polyamine binding proteins or replaced residues of the spermidine binding site, including the critical tryptophan 271, which governs substrate specificity for spermidine compared to putrescine [45]. The more common sequence variants in particular clustered in the domain that determines the binding specificity for biogenic amines. Considering the fact that the SpuE-mediated uptake of spermidine from the mammalian host leads to transcriptional upregulation of the type III secretion system [45, 46], the major virulence determinant of *P. aeruginosa*, it is reasonable to assume that within the scenario of host–pathogen interactions the sequence diversity of SpuE is translated into a batch of functional variants of differential pathogenicity.

LasR and *RhlR*

Two interlinked *N*-acyl homoserine lactone (AHL)-dependent regulatory circuits contribute to quorum sensing in *P. aeruginosa*, i.e. the *las* system, which consists of the transcriptional activator LasR and the AHL synthase LasI directing the synthesis of *N*-3-oxo-dodecanoyl-homoserine lactone (3-oxo-C_{12}-HSL) and the *rhl* system, which consists of the transcriptional regulator RhlR and the AHL synthase RhlI directing the synthesis of *N*-butanoyl-homoserine lactone (C_4-HSL) [47]. Both regulators LasR and RhlR had been identified as hot spots of mutation in our collection of serial isolates from CF airways [21]. Consistent with the high d_N/d_S ratio described above, we only observed 26 synonymous nucleotide substitutions, but 63 amino acid sequence variants and 15 stop mutations in the *lasR* genes of our strain panel (Table S4). A high frequency of *lasR* missense and nonsense mutations has already been reported for CF strain collections from the USA [12, 23, 25] and Scandinavia [24], but in contrast to the literature [12, 25] we identified a similarly

Table 6. Distribution of amino acid replacements in the *P. aeruginosa* strain panel

Affected amino acid	No. of missense mutations present			Average Dayhoff score of missense mutations		
	CF and non-CF	CF	non-CF	CF and non-CF	CF	non-CF isolates
Alanine	60	37	21	50	47	46
Arginine	13	28	12	8	5	4
Asparagine	9	9	1	41	37	43
Aspartic acid	16	28	7	42	30	54
Cysteine	1	7	1	3	7	1
Glutamine	3	14	6	21	6	11
Glutamic acid	15	11	7	45	21	34
Glycine	8	25	14	45	25	26
Histidine	6	6	5	9	13	8
Isoleucine	11	12	4	41	22	13
Leucine	10	22	3	14	9	17
Lysine	4	7	2	36	43	29
Methionine	6	5	0	4	8	
Phenylalanine	2	4	4	11	17	9
Proline	12	21	11	23	15	14
Serine	13	13	5	41	32	24
Threonine	18	18	6	30	45	17
Tryptophan	1	8	4	0	1	1
Tyrosine	2	9	4	13	2	2
Valine	20	25	15	41	27	34

high mutation rate of 30% in the isolates from acute infections and patients with COPD (Table S4). *LasR* missense mutations were only rare in isolates from the environment (3 %). Our genotyping data of a geographically diverse strain collection demonstrate that non-synonymous *lasR* mutations occur with similar frequency in CF, COPD and acute human infections.

Next, to increase the statistical power of a more refined analysis we compiled the data on *lasR* missense variants of 197 strains from the Seattle [25], Copenhagen [24] and our collections. Benign and non-conservative amino acid substitutions had the same share of ~40 and ~60% in the three strain collections and in both CF and non-CF isolates. The pronounced diversity of the 124 missense variants showed up by the large portion of unique mutations (71%) and by multiple amino acid substitutions at 28 residues. Missense mutations clustered in the C-terminal DNA-binding domain (P=10^{-11}) (Fig. 3a). Missense mutation rates per amino acid were 1.8 in the DNA-binding domain, 0.9 in the AHL-binding site [48] and 0.4 elsewhere in the protein. The highest mutation rate of 1.9 was observed at the 16 positions that are obligatorily conserved among LasR orthologues (P=0.0004) [49]. The preponderance of mutations at the residues that are evolutionarily most conserved and/or instrumental for AHL or DNA binding strongly indicates that most missense mutations modulate or – like the numerous nonsense mutations – abrogate the natural function of LasR.

The mutation spectrum in *rhlR* was similar to that in *lasR*, albeit the frequency of mutations was lower and stop mutations were missing (Table S4). The vast majority of the 24 amino acid substitutions were non-conservative and present in only 1 or 2 strains. Missense variants were identified in 10–15% of the isolates from CF, COPD and acute infections, but not in any strain of environmental origin. In summary, the population genetics of the quorum sensing regulators *lasR* and *rhlR* is characterized by ubiquitous synonymous SNVs and a vast spectrum of rare non-neutral missense mutations in human habitats.

AlgU and *AlgG*

The onset of chronic colonization of CF lungs is characterized by the emergence of alginate-overproducing mucoid morphotypes of *P. aeruginosa* [50], although after 10 or more years of colonization most mucoid clones have reverted to a non-mucoid phenotype [21]. The transcriptional programme for alginate production is induced by the extracytoplasmic sigma

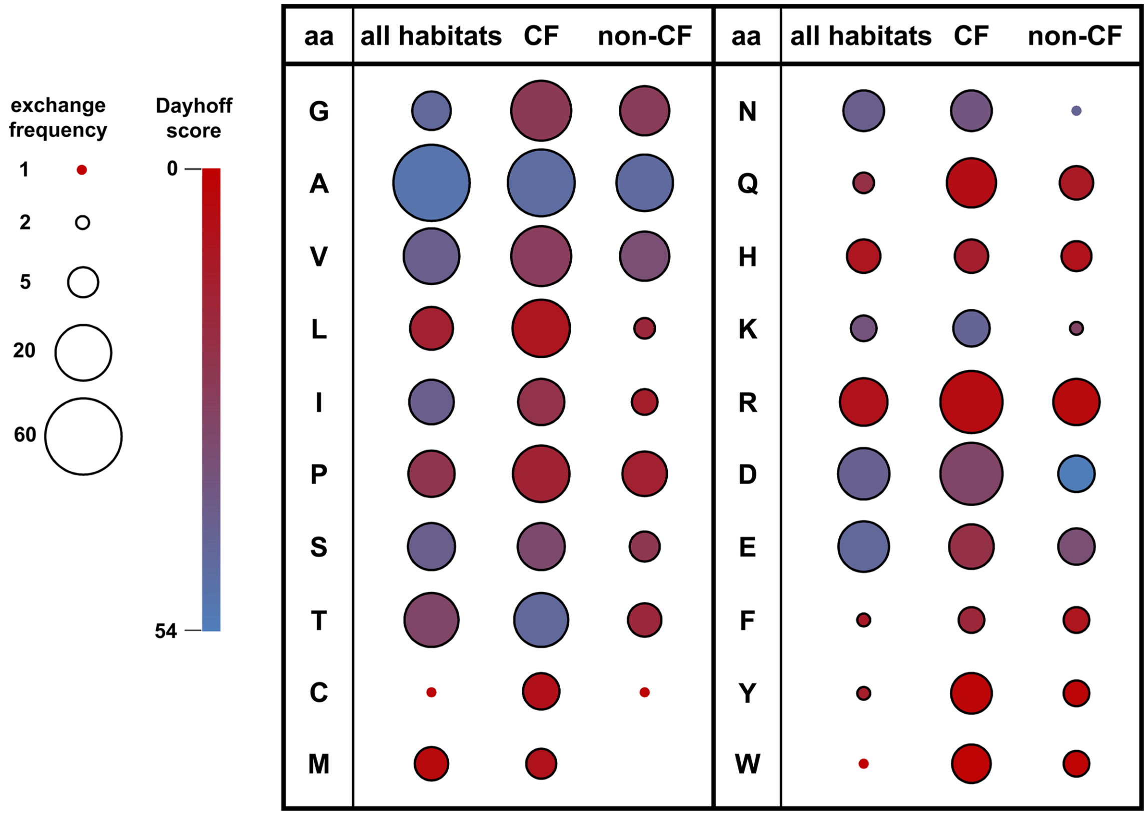

Fig. 1. Distribution of amino acid exchanges. The figure summarizes the frequency and potential functional severity of amino acid exchanges observed in the *P. aeruginosa* strain panel. The observed numbers of events affecting a certain amino acid either in strains from all habitats, exclusively in CF isolates or in non-CF isolates only are represented by the diameters of the respective circles. Please note that the diameters are adjusted to a logarithmic scale, only the spots for frequencies values of 1 are displayed in an estimated size and without a black outer ring. The colour of the circles represents the average Dayhoff score of all exchanges affecting the respective amino acid. Dayhoff scores are taken from a matrix containing frequencies of observed amino acid exchanges in sequences of functionally equivalent proteins. Thus they are taken as probability values for potential functional conservation or changes introduced by single exchanges, with low values indicating potentially severe functional changes and higher values hinting at likely less drastic or even very little effect on protein function.

factor AlgU if the repression by the cognate anti-sigma factor MucA is removed [51]. The sequencing of *algU* in our strain panel revealed 39 synonymous and 24 non-synonymous SNVs (Table S4). Sequence variants were significantly more frequent among the CF isolates (P_{corr}=4×10^{-6}). The *algU* missense mutations were evenly distributed in the gene without any predilection for sequence motifs or positions evolutionarily conserved in orthologues. The mutation spectrum, however, was different from that of a collection of non-mucoid revertants of *mucA* mutants that carried significantly more nonsense and frameshift mutations in the *algU* gene (*P*=0.009) [22, 26].

P. aeruginosa alginate is an unbranched anionic polysaccharide that consists of blocks of partially acetylated homopolymeric mannuronic acid and of heteropolymeric disaccharide blocks of mannuronic acid and its C5 epimer, α-L-guluronate [52]. The epimerization of mannuronic to guluronic acid within the mannuroate polymer is catalyzed by the periplasmic alginate epimerase AlgG [53]. Amplicon sequencing identified 66 synonymous and 71 non-synonymous SNVs, with a codon being targeted by 2 missense substitutions 9 times (Table S4). Of the 23 missense polymorphisms with a prevalence of more than 1%, 8 mutants were only present in isolates from CF airways (*P*=0.005). Coding sequence variants were not evenly distributed in the primary AlgG sequence (*P*=0.033) (Fig. 3b). They clustered in seven 6–22 amino acid long stretches, with the frequent polymorphisms present in all habitats primarily being located in the N-terminal tail lacking regular secondary structure [54]. All other clusters of point mutations resided within the coils of the parallel β-helix fold [54] and were populated by missense variants exclusively detected in CF isolates. According to the resolved three-dimensional structure of AlgG of *Pseudomonas syringae* [54] and secondary structure predictions for AlgG of *P. aeruginosa* [55] the parallel β-helix of AlgG comprises 12 coils, each of which is made up of 3 β-sheets, PB1, PB2 and PB3, which are linked by turns. Mutations in PB1 accounted for more than half of the missense variants in the coils. Conversely, apart from neutral

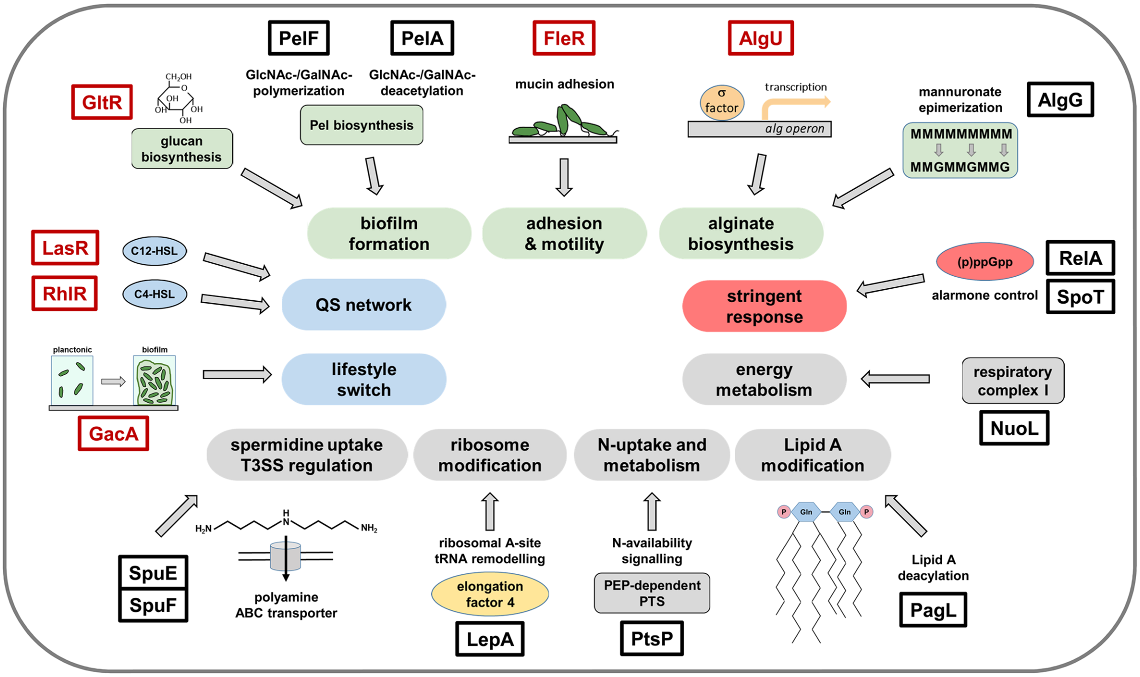

Fig. 2. Overview on mutation hot spot loci and the corresponding functions. The figure displays the proteins encoded in the loci found to be most frequently mutated in *P. aeruginosa* strains from CF background, their individual traits and functional category. Please note that only 17 of the 19 proteins are shown here, as for 2 frequently mutated loci (PA4391 and PA5048) functional data is still lacking, and the encoded proteins are still classified as hypothetical in the databases.

polymorphisms at position 127, the canonical AlgG sequence was 100% conserved in its five helices, indicating that the protein can incorporate novel amino acids in the coils, but does not tolerate any change in the helices (Fig. 3b). Ten and seven missense mutants were located within or close to the active site region or the polymannuronate-binding site, respectively (Fig. 3b). Epimerase activity and/or alginate secretion should be affected in these missense mutants. In summary, the carbohydrate-binding/sugar hydrolysis domain that is made up by coils 4–10 [54, 55] is the target of frequent mutations in *P. aeruginosa* isolates from the CF lung, but not from other habitats.

PelA and *PelF*

The seven-gene *pel* operon encodes the proteins for the biosynthesis, processing, export and degradation of the Pel exopolysaccharide, which is an important constituent of the biofilm matrix of *P. aeruginosa*. Pel is a linear cationic exopolysaccharide containing 1→4 glycosidic linkages of partially deacetylated N-acetylgalactosamine and N-acetylglucosamine [56]. PelF catalyzes the biosynthesis of the amino sugar polymer [56] and PelA then partially removes the acetyl groups from GalNAc and GlcNAc [57]. PelA is a multidomain bifunctional enzyme [57] that is involved in both the formation and dispersion of Pel biofilms via its deacetylase and glycoside hydrolase domains, respectively [58]. Amplicon sequencing of *pelF* identified 116 synonymous substitutions and 62 missense mutations at 55 codons (Table S4). The missense variants were non-evenly distributed in the primary sequence (P=0.02) and were preferentially located at evolutionarily not conserved positions (P=3×10^{-5}). The vast majority of amino acid replacements should not impair protein function, as indicated by their high average Dayhoff index of 38 and the frequent change to the canonical amino acid of the closest orthologue PelF of *Pseudomonas protegens*. The *pelA* gene carried the highest number of synonymous (179) and non-synonymous (118) SNVs among the 19 genes (Table S4) and was the only gene with a higher frequency of missense mutations in *P. aeruginosa* strains from non-CF habitats than in isolates from CF airways (P=0.02) (Table 5). Amino acid sequence variants were most prevalent in both abundance and mutation frequency in the N-terminal TAT signal sequence followed by the glycoside hydrolase domain and least common in the deacetylase domain [58] (Fig. 3c). The broad distribution of 43 common amino acid replacements found in 10 or more strains from diverse habitats demonstrates that a large repertoire of PelA protein variants is propagated in the *P. aeruginosa* population.

PagL

An opposite behaviour was observed for the population genetics of *pagL* that encodes the lipopolysaccharide lipid A 3-O-deacylase. Besides six probably neutral changes to the amino acid present in *pagL* homologues [59], we predominantly identified mutations in singular CF isolates, i.e. nine missense mutations at conserved sites, seven nonsense

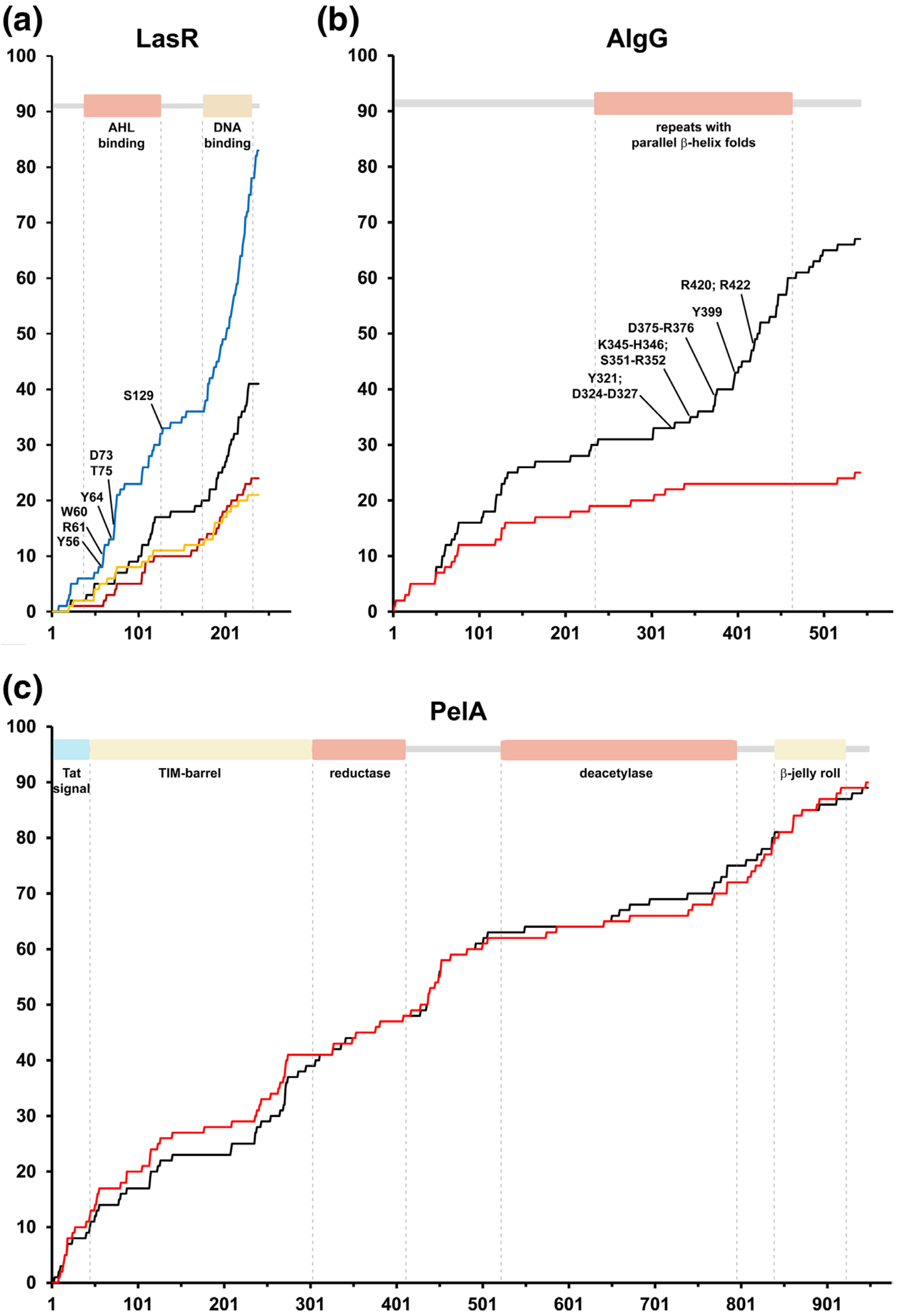

Fig. 3. Amino acid sequence variants in (a) LasR, (b) AlgG and (c) PelA. The cumulative number of mutations causing amino acid exchanges is plotted versus the position in the primary amino acid sequence from the N- to the C-terminus. Each exchange is counted once independent of its detection in single or several strains. Divergent exchanges observed at the same position are counted as independent events. For each protein, the observed mutations are summarized individually for the subsets of the 345 CF isolates (black line) and the 189 non-CF isolates (red line). In case of LasR (a), mutations detected in other studies on *P. aeruginosa* CF isolates are also displayed (blue line, Seattle, USA [25]; yellow line, Copenhagen, Denmark [24]). Known functionally important positions and domains within the protein sequences are indicated. For LasR, the DNA-binding region and the major amino acids in the AHL-binding pocket are shown [48]. Similarly, the polymannuronate embedding region containing 24 aa repeats with β-helix folds is shown for AlgG (b). The indicated amino acids designate the major residues of the AlgG catalytic centre and substrate-binding sites [54, 55]. For PelA (c), the structural domains are shown [57].

mutations and one loss of the start codon (Table S4). PagL releases the primary hydroxyl fatty acid moiety at the 3 position of lipid A and thereby converts the hexa-acylated form of lipid A to the less hydrophobic penta-acylated form [59, 60], which facilitates evasion from the mammalian host response and confers resistance to the peptide antibiotic polymyxin [61]. This dual role of PagL in resistance and immune evasion makes it plausible why all but one of the PagL rare missense mutants were exclusively detected in isolates from CF lungs.

Executive summary and conclusion

This report provides information about the molecular epidemiology of 19 *P. aeruginosa* genes. These loci represent the hot spots of microevolution in the airways of the six most mildly and six most severely affected CF patients seen at the CF clinic Hannover who had become chronically colonized with *P. aeruginosa* in the 1980s [21]. Amplicon sequencing of a large strain collection indeed revealed a significantly larger portion of sequence variants in the CF isolates, but this overrepresentation was mainly due to the larger share of synonymous substitutions. These silent mutations may potentially influence the utilization of the tRNA pool during translation, but we consider it more likely that the mutation rates may be higher in CF lungs than in other habitats and that the silent mutations are tolerated, whereas most non-synonymous substitutions are rapidly eliminated from the bacterial community. This interpretation is in line with the d_N/d_S ratios of the genes that are below the average value of the genes in the *P. aeruginosa* genome [62, 63], indicating strong purifying selection for this gene panel.

CF strains showed higher relative and absolute d_N/d_S values and higher frequencies of amino acid sequence variants primarily in targets such as PagL, GacA, AlgU and AlgG that shape the phenotypic conversion of *P. aeruginosa* in CF lungs and may be considered as potential drug targets (Table 7) [64].

Table 7. Adaptation of *P. aeruginosa* to the CF environment

Gene*	Non-conservative coding variants in the CF habitat			
	Detection	CF/non-CF skew	Positive selection	Potential drug target
algG	Yes	Yes		Yes
algU	Yes	Yes	Yes	
gacA	Yes			
lasR	Yes		Yes	Yes
pagL	Yes	Yes	Yes	
pelA	Yes			
rhlR	Yes		Yes	
spoT	Yes	Yes		
spuE	Yes			Yes

*Loci are not shown that lacked any coding variant at conserved positions or any missense mutation with low Dayhoff score.

Most coding variants were only identified in single CF isolates, indicating that they are rare and/or *de novo* mutants confined to a single or a few CF hosts. The exception to this rule was the epimerase AlgG. AlgG missense variants were broadly distributed among CF isolates. AlgG activity determines the composition of mannuorate–mannuorate and mannuorate–guluronate blocks in the alginate polymer, which together with the subsequent acetylation may fine tune the structure, charge and rheological properties of the anionic exopolysaccharide and thus even the global features of the biofilm in the CF airways. Co-colonization with multiple AlgG variants could confer a gain of fitness to the *P. aeruginosa* community to thrive in the atypical CF lung habitat.

Likewise numerous variants were characteristic for PelA, which is instrumental for the production and dispersion of the cationic Pel exopolysaccharide. PelA showed the largest diversity among the isolates from inanimate aquatic habitats. Like AlgG variants in CF lungs, the PelA variants may endow competitive fitness to the *P. aeruginosa* community to thrive in the environment.

Lastly, we noted with surprise that LasR and RhlR quorum sensing variants including loss-of-function *lasR* mutants are as frequent in acute human infections as in chronic CF lung infections [23–25]. Consistent with the knowledge that the RhlR quorum sensing circuit is often functional (and functioning) in *lasR* loss-of-function mutants [25, 65], we did not observe any stop mutation in the *rhlR* gene in our strain collection. The rewiring of the quorum sensing circuitry seems to be a general phenomenon in human infections with *P. aeruginosa*.

Funding information
This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB900/3 – 158989968 – A2 and Z1.

Acknowledgements
The provision of strains by members of the *Pseudomonas* community is gratefully acknowledged. The authors thank Nina Cramer and Samira Mielke for helpful discussions and practical support during the initiation period of the study.

Author contributions
S.F.: conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing – review and editing. J.K.: conceptualization, data curation, investigation, methodology, validation, visualization, writing – review and editing. M.G.S.: investigation, validation. M.D.: investigation. L.W.: methodology, resources. B.T.: conceptualization, formal analysis, funding acquisition, investigation, methodology, writing – original draft, writing – review and editing.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nat Rev Genet* 2015;16:45–56.
2. Stoltz DA, Meyerholz DK, Welsh MJ. Origins of cystic fibrosis lung disease. *N Engl J Med* 2015;372:351–362.
3. Ratjen F, Bell SC, Rowe SM, Goss CH, Quittner AL, *et al*. Cystic fibrosis. *Nat Rev Dis Primers* 2015;1:15010.
4. Elborn JS. Cystic fibrosis. *Lancet* 2016;388:2519–2531.

5. Blanchard AC, Waters VJ. Microbiology of cystic fibrosis airway disease. *Semin Respir Crit Care Med* 2019;40:727–736.
6. Cramer N, Wiehlmann L, Tümmler B. Clonal epidemiology of *Pseudomonas aeruginosa* in cystic fibrosis. *Int J Med Microbiol* 2010;300:526–533.
7. Parkins MD, Somayaji R, Waters VJ. Epidemiology, biology, and impact of clonal *Pseudomonas aeruginosa* infections in cystic fibrosis. *Clin Microbiol Rev* 2018;31:e00019-18.
8. Döring G, Flume P, Heijerman H, Elborn JS, Consensus Study Group. Treatment of lung infection in patients with cystic fibrosis: current and future strategies. *J Cyst Fibros* 2012;11:461–479.
9. Taccetti G, Francalanci M, Pizzamiglio G, Messore B, Carnovale V, *et al*. Cystic Fibrosis: Recent insights into inhaled antibiotic treatment and future perspectives. *Antibiotics (Basel)* 2021;10:338.
10. Rossi E, La Rosa R, Bartell JA, Marvig RL, Haagensen JAJ, *et al*. *Pseudomonas aeruginosa* adaptation and evolution in patients with cystic fibrosis. *Nat Rev Microbiol* 2021;19:331–342.
11. Camus L, Vandenesch F, Moreau K. From genotype to phenotype: adaptations of *Pseudomonas aeruginosa* to the cystic fibrosis environment. *Microb Genom* 2021;7.
12. Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, *et al*. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A* 2006;103:8487–8492.
13. Cramer N, Klockgether J, Wrasman K, Schmidt M, Davenport CF, *et al*. Microevolution of the major common *Pseudomonas aeruginosa* clones C and PA14 in cystic fibrosis lungs. *Environ Microbiol* 2011;13:1690–1704.
14. Feliziani S, Marvig RL, Luján AM, Moyano AJ, Di Rienzo JA, *et al*. Coexistence and within-host evolution of diversified lineages of hypermutable *Pseudomonas aeruginosa* in long-term cystic fibrosis infections. *PLoS Genet* 2014;10:e1004651.
15. van Mansfeld R, de Been M, Paganelli F, Yang L, Bonten M, *et al*. Within-host evolution of the Dutch high-prevalent *Pseudomonas aeruginosa* clone ST406 during chronic colonization of a patient with cystic fibrosis. *PLoS One* 2016;11:e0158106.
16. Bianconi I, D'Arcangelo S, Esposito A, Benedet M, Piffer E, *et al*. Persistence and microevolution of *Pseudomonas aeruginosa* in the cystic fibrosis lung: A single-patient longitudinal genomic study. *Front Microbiol* 2019;9:3242.
17. Yang L, Jelsbak L, Marvig RL, Damkiær S, Workman CT, *et al*. Evolutionary dynamics of bacteria in a human host environment. *Proc Natl Acad Sci U S A* 2011;108:7481–7486.
18. 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;9:e1003741.
19. Marvig RL, Sommer LM, Molin S, Johansen HK. Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. *Nat Genet* 2015;47:57–64.
20. Marvig RL, Dolce D, Sommer LM, Petersen B, Ciofu O, *et al*. Within-host microevolution of *Pseudomonas aeruginosa* in Italian cystic fibrosis patients. *BMC Microbiol* 2015;15:218.
21. Klockgether J, Cramer N, Fischer S, Wiehlmann L, Tümmler B. Long-term microevolution of *Pseudomonas aeruginosa* differs between mildly and severely affected cystic fibrosis lungs. *Am J Respir Cell Mol Biol* 2018;59:246–256.
22. Ciofu O, Lee B, Johannesson M, Hermansen NO, Meyer P, *et al*. Investigation of the algT operon sequence in mucoid and non-mucoid *Pseudomonas aeruginosa* isolates from 115 Scandinavian patients with cystic fibrosis and in 88 in vitro non-mucoid revertants. *Microbiology (Reading)* 2008;154:103–113.
23. Hoffman LR, Kulasekara HD, Emerson J, Houston LS, Burns JL, *et al*. *Pseudomonas aeruginosa lasR* mutants are associated with cystic fibrosis lung disease progression. *J Cyst Fibros* 2009;8:66–70.
24. Bjarnsholt T, Jensen PØ, Jakobsen TH, Phipps R, Nielsen AK, *et al*. Quorum sensing and virulence of *Pseudomonas aeruginosa* during lung infection of cystic fibrosis patients. *PLoS One* 2010;5:e10115.
25. Feltner JB, Wolter DJ, Pope CE, Groleau M-C, Smalley NE, *et al*. LasR variant cystic fibrosis isolates reveal an adaptable quorum-sensing hierarchy in *Pseudomonas aeruginosa*. *mBio* 2016;7:e01513-16.
26. Candido Caçador N, Paulino da Costa Capizzani C, Gomes Monteiro Marin Torres LA, Galetti R, Ciofu O, *et al*. Adaptation of *Pseudomonas aeruginosa* to the chronic phenotype by mutations in the algTmucABD operon in isolates from Brazilian cystic fibrosis patients. *PLoS One* 2018;13:e0208013.
27. Wiehlmann L, Cramer N, Tümmler B. Habitat-associated skew of clone abundance in the *Pseudomonas aeruginosa* population. *Environ Microbiol Rep* 2015;7:955–960.
28. Fischer S, Dethlefsen S, Klockgether J, Tümmler B. Phenotypic and genomic comparison of the two most common ExoU-positive *Pseudomonas aeruginosa* clones, PA14 and ST235. *mSystems* 2020;5:e01007-20.
29. Wiehlmann L, Wagner G, Cramer N, Siebert B, Gudowius P, *et al*. Population structure of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 2007;104:8101–8106.
30. Dale JW, Greenaway PJ. Preparation of chromosomal DNA from *E. coli*. *Methods Mol Biol* 1985;2:197–200.
31. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* 2914:btu170.
32. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013.
33. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, *et al*. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012;6:80–92.
34. Cramer N, Wiehlmann L, Ciofu O, Tamm S, Høiby N, *et al*. Molecular epidemiology of chronic *Pseudomonas aeruginosa* airway infections in cystic fibrosis. *PLoS One* 2012;7:e50731.
35. Schurr MJ, Martin DW, Mudd MH, Deretic V. Gene cluster controlling conversion to alginate-overproducing phenotype in *Pseudomonas aeruginosa*: functional analysis in a heterologous host and role in the instability of mucoidy. *J Bacteriol* 1994;176:3375–3382.
36. Williams P, Cámara M. Quorum sensing and environmental adaptation in *Pseudomonas aeruginosa*: a tale of regulatory networks and multifunctional signal molecules. *Curr Opin Microbiol* 2009;12:182–191.
37. Schuster M, Sexton DJ, Diggle SP, Greenberg EP. Acyl-homoserine lactone quorum sensing: from evolution to application. *Annu Rev Microbiol* 2013;67:43–63.
38. Smith EE, Sims EH, Spencer DH, Kaul R, Olson MV. Evidence for diversifying selection at the pyoverdine locus of *Pseudomonas aeruginosa*. *J Bacteriol* 2005;187:2138–2147.
39. Ganne G, Brillet K, Basta B, Roche B, Hoegy F, *et al*. Iron release from the siderophore pyoverdine in *Pseudomonas aeruginosa* involves three new actors: FpvC, FpvG, and FpvH. *ACS Chem Biol* 2017;12:1056–1065.
40. Dayhoff MO. *Atlas of Protein Sequence and Structure*. Washington, DC: National Biomedical Research Foundation; 1978.
41. Pausch P, Abdelshahid M, Steinchen W, Schäfer H, Gratani FL, *et al*. Structural basis for regulation of the opposing (p)ppGpp synthetase and hydrolase within the stringent response orchestrator Rel. *Cell Rep* 2020;32:108157.
42. Arenz S, Abdelshahid M, Sohmen D, Payoe R, Starosta AL, *et al*. The stringent factor RelA adopts an open conformation on the ribosome to stimulate ppGpp synthesis. *Nucleic Acids Res* 2016;44:6471–6481.
43. Winther KS, Roghanian M, Gerdes K. Activation of the stringent response by loading of RelA-tRNA complexes at the ribosomal A-site. *Molecular Cell* 2018;70:95-105.
44. Pletzer D, Blimkie TM, Wolfmeier H, Li Y, Baghela A, *et al*. The stringent stress response controls proteases and global regulators under optimal growth conditions in *Pseudomonas aeruginosa*. *mSystems* 2020;5:e00495-20.

45. Wu D, Lim SC, Dong Y, Wu J, Tao F, *et al*. Structural basis of substrate binding specificity revealed by the crystal structures of polyamine receptors SpuD and SpuE from *Pseudomonas aeruginosa*. *J Mol Biol* 2012;416:697–712.
46. Zhang Y, Sun X, Qian Y, Yi H, Song K, *et al*. A potent anti-SpuE antibody allosterically inhibits Type III secretion system and attenuates virulence of *Pseudomonas aeruginosa*. *J Mol Biol* 2019;431:4882–4896.
47. Juhas M, Eberl L, Tümmler B. Quorum sensing: the power of cooperation in the world of *Pseudomonas*. *Environ Microbiol* 2005;7:459–471.
48. Bottomley MJ, Muraglia E, Bazzo R, Carfì A. Molecular insights into quorum sensing in the human pathogen *Pseudomonas aeruginosa* from the structure of the virulence regulator LasR bound to its autoinducer. *J Biol Chem* 2007;282:13592–13600.
49. Qiu H, Li Y, Dai W. Codon-usage frequency mediated SNPs selection in *lasR* gene of cystic fibrosis *Pseudomonas aeruginosa* isolates. *Microbiol Res* 2019:137–143.
50. Govan JR, Deretic V. Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiol Rev* 1996;60:539–574.
51. Damron FH, Goldberg JB. Proteolytic regulation of alginate overproduction in *Pseudomonas aeruginosa*. *Mol Microbiol* 2012;84:595–607.
52. Sherbrock-Cox V, Russell NJ, Gacesa P. The purification and chemical characterisation of the alginate present in extracellular material produced by mucoid strains of *Pseudomonas aeruginosa*. *Carbohydr Res* 1984;135:147–154.
53. Franklin MJ, Chitnis CE, Gacesa P, Sonesson A, White DC, *et al*. *Pseudomonas aeruginosa* AlgG is a polymer level alginate C5-mannuronan epimerase. *J Bacteriol* 1994;176:1821–1830.
54. Wolfram F, Kitova EN, Robinson H, Walvoort MTC, Codée JDC, *et al*. Catalytic mechanism and mode of action of the periplasmic alginate epimerase AlgG. *J Biol Chem* 2014;289:6006–6019.
55. Douthit SA, Dlakic M, Ohman DE, Franklin MJ. Epimerase active domain of *Pseudomonas aeruginosa* AlgG, a protein that contains a right-handed beta-helix. *J Bacteriol* 2005;187:4573–4583.
56. Jennings LK, Storek KM, Ledvina HE, Coulon C, Marmont LS, *et al*. Pel is a cationic exopolysaccharide that cross-links extracellular DNA in the *Pseudomonas aeruginosa* biofilm matrix. *Proc Natl Acad Sci U S A* 2015;112:11353–11358.
57. Colvin KM, Alnabelseya N, Baker P, Whitney JC, Howell PL, *et al*. PelA deacetylase activity is required for Pel polysaccharide synthesis in *Pseudomonas aeruginosa*. *J Bacteriol* 2013;195:2329–2339.
58. Cherny KE, Sauer K. Untethering and degradation of the polysaccharide matrix are essential steps in the dispersion response of *Pseudomonas aeruginosa* biofilms. *J Bacteriol* 2020;202.
59. Geurtsen J, Steeghs L, Hove JT, van der Ley P, Tommassen J. Dissemination of lipid A deacylases (pagL) among gram-negative bacteria: identification of active-site histidine and serine residues. *J Biol Chem* 2005;280:8248–8259.
60. Rutten L, Geurtsen J, Lambert W, Smolenaers JJM, Bonvin AM, *et al*. Crystal structure and catalytic mechanism of the LPS 3-O-deacylase PagL from *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 2006;103:7071–7076.
61. Han M-L, Velkov T, Zhu Y, Roberts KD, Le Brun AP, *et al*. Polymyxin-induced Lipid A deacylation in *Pseudomonas aeruginosa* perturbs polymyxin penetration and confers high-level resistance. *ACS Chem Biol* 2018;13:121–130.
62. Hilker R, Munder A, Klockgether J, Losada PM, Chouvarine P, *et al*. Interclonal gradient of virulence in the *Pseudomonas aeruginosa* pangenome from disease and environment. *Environ Microbiol* 2015;17:29–46.
63. Fischer S, Klockgether J, Morán Losada P, Chouvarine P, Cramer N, *et al*. Intraclonal genome diversity of the major *Pseudomonas aeruginosa* clones C and PA14. *Environ Microbiol Rep* 2016;8:227–234.
64. Tümmler B. Clonal variations in *Pseudomonas aeruginosa*. In: Ramos JL and Levesque RC (eds). *Pseudomonas. Volume 4. Molecular Biology of Emerging Issues*. New York: Springer; 2006. pp. 35–68.
65. Kostylev M, Kim DY, Smalley NE, Salukhe I, Greenberg EP, *et al*. Evolution of the *Pseudomonas aeruginosa* quorum-sensing hierarchy. *Proc Natl Acad Sci U S A* 2019;116:7027–7032.