

## RESEARCH ARTICLE

Genome-wide association study of signature genetic alterations among *Pseudomonas aeruginosa* cystic fibrosis isolatesWontae Hwang<sup>1,2</sup>, Ji Hyun Yong<sup>1,2</sup>, Kyung Bae Min<sup>1,2</sup>, Kang-Mu Lee<sup>1,2</sup>, Ben Pascoe<sup>3</sup>, Samuel K Sheppard<sup>3</sup>, Sang Sun Yoon<sup>1,2,4,5\*</sup>

**1** Department of Microbiology and Immunology, Seoul, Republic of Korea, **2** Brain Korea 21 PLUS Project for Medical Sciences, Seoul, Republic of Korea, **3** The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, United Kingdom, **4** Institute for Immunology and Immunological Diseases, Seoul, Republic of Korea, **5** Severance Biomedical Science Institute, Yonsei University College of Medicine, Seoul, Republic of Korea

\* [sangsun\\_yoon@yuhs.ac](mailto:sangsun_yoon@yuhs.ac)

## OPEN ACCESS

**Citation:** Hwang W, Yong JH, Min KB, Lee K-M, Pascoe B, Sheppard SK, et al. (2021) Genome-wide association study of signature genetic alterations among *Pseudomonas aeruginosa* cystic fibrosis isolates. *PLoS Pathog* 17(6): e1009681. <https://doi.org/10.1371/journal.ppat.1009681>

**Editor:** Alice Prince, Columbia University, UNITED STATES

**Received:** April 6, 2021

**Accepted:** May 31, 2021

**Published:** June 23, 2021

**Copyright:** © 2021 Hwang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its [Supporting information](#) files.

**Funding:** This work was supported by grants received by S.S.Y. from the National Research Foundation (NRF) of Korea, which is funded by the Korean Government (2017M3A9F3041233 and 2019R1A6A1A03032869). This research was also supported by a grant received by S.S.Y. from the Korean Health Technology R&D Project through the Korean Health Industry Development Institute

## Abstract

*Pseudomonas aeruginosa* (PA) is an opportunistic pathogen that causes diverse human infections including chronic airway infection in patients with cystic fibrosis (CF). Comparing the genomes of CF and non-CF PA isolates has great potential to identify the genetic basis of pathogenicity. To gain a deeper understanding of PA adaptation in CF airways, we performed a genome-wide association study (GWAS) on 1,001 PA genomes. Genetic variations identified among CF isolates were categorized into (i) alterations in protein-coding regions, either large- or small-scale, and (ii) polymorphic variation in intergenic regions. We introduced each CF-associated genetic alteration into the genome of PAO1, a prototype PA strain, and validated the outcomes experimentally. Loci readily mutated among CF isolates included genes encoding a probable sulfatase, a probable TonB-dependent receptor (PA2332~PA2336), L-cystine transporter (YecS, PA0313), and a probable transcriptional regulator (PA5438). A promoter region of a heme/hemoglobin uptake outer membrane receptor (PhuR, PA4710) was also different between the CF and non-CF isolate groups. Our analysis highlights ways in which the PA genome evolves to survive and persist within the context of chronic CF infection.

## Author summary

Understanding how human pathogens, such as *Pseudomonas aeruginosa* (PA), adapt to the host environment during prolonged infections can provide important information for the treatment of disease. In this study, we compared the genome sequences of PA isolates from cystic fibrosis (CF) and non-CF patients and identified genetic variation that was significantly associated with CF disease. CF-associated genetic variations included: (i) the deletion of three amino acids in a transcriptional regulator (PA5438), which leads to decreased antibiotic susceptibility, enhanced virulence expression and sustained biofilm production in the presence of transferrin, and (ii) the insertion of three amino acids in a

(KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HI14C1324). All funders listed had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

cystine transporter (PA0313), which results in decreased pyochelin production. Furthermore, single nucleotide replacement in a promoter was identified and shown to upregulate the cellular process for iron acquisition. Our bioinformatics analyses, linked with experimental validation via genetic manipulation and phenotypic investigation of mutants, provides novel insights into the genomic plasticity of PA during chronic CF infection.

## Introduction

*Pseudomonas aeruginosa* (PA) is an ubiquitous gram-negative bacterium that can cause disease in plants and animals [1]. PA infection can cause acute syndromes such as pneumonia and bloodstream infections, as well as chronic airway infections in patients with cystic fibrosis (CF). CF is a well-known genetic disorder caused by a mutated cystic fibrosis transmembrane conductance regulator (CFTR) protein. CFTR disruption alters the condition of the lung such that the increasingly dehydrated viscous mucus layer provides a favorable habitat for several pathogens, such as PA, which can infect opportunistically and persist over an extended period of time [2].

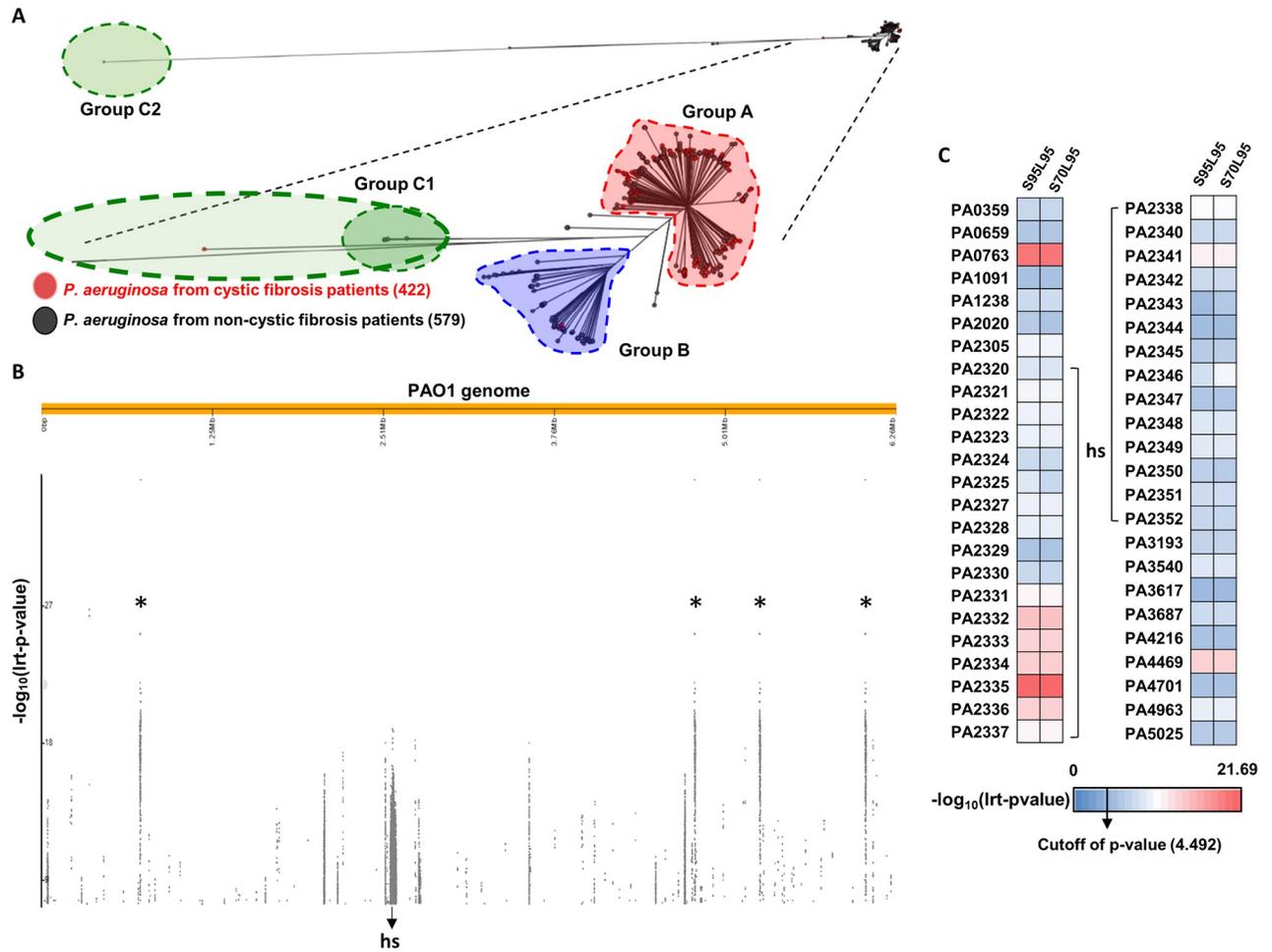
The ecological versatility of PA is thought to be associated with its relatively large genome containing numerous regulatory genes that confer an advantage in adapting during prolonged infections [3]. Within-host mutation of PA isolates from a single CF patient have been analyzed to better understand the underlying genetics of adaptation [4]. Expression studies have also been performed on PA grown in CF sputum and have identified altered expression of genes encoding amino acid biosynthesis and degradation, and quinolone signaling [5,6]. The *lasR* gene, which encodes an important quorum-sensing regulator, has frequently been detected among isolates from CF patients [7,8], however, the significance of this mutation is unclear. Mutations leading to increased antibiotic tolerance have also been identified in PA inhabiting the CF airway [9–11]. Furthermore, PA clonal lineages vary across different CF patients, depending on the sputum composition [12–15], and even within an individual patient over time [16,17]. Together, these findings suggest that PA is an adaptable organism that responds flexibly to changing environments.

Most large-scale genetic studies have focused principally on CF isolates. Therefore, in the absence of a comparator non-CF PA population, it is difficult to identify specific genetic variations that are associated with the CF environment. In this study, we compared 1,001 PA genomes sampled from both CF and non-CF clinical isolates. Genome-wide association studies (GWASs) have been useful in uncovering causal relationships between genetic variations and disease phenotypes in human populations [18], with over 4,000 human GWAS conducted to date [19]. In contrast to human genomes, bacterial genomes of the same species can vary in size, gene repertoire, and gene arrangement [20]. Because of this genomic variability, additional considerations must be made when performing bacterial GWAS [21–25]. In this study, we selected the genomes of PA isolates of known origin and performed a GWAS based on k-mer counting [22,23,26], a modified method that permits association mapping of genes and intergenic regions of the PAO1 reference genome. Results provided here expand our current understanding of how gene-level changes correlate with mechanisms of PA adaptation to the CF lung environment.

## Results and discussion

### *P. aeruginosa* population structure

Most studies aimed at better understanding chronic PA infection in CF patients have focused only on CF isolates [14,16]. It is therefore unclear whether genetic variations highlighted in



**Fig 1. Visualization of significantly different regions and proteins with large-scale mutation.** (A) Phylogenetic tree was constructed with 1,001 genomes with known host disease information. Black and red leaves each indicate non-CF and CF isolates, and the numbers of CF and non-CF genomes for constructing the phylogenetic tree are placed inside brackets. Most of the CF isolates are located in group A, which includes the majority of genomes, and the smaller groups are designated group B, C1 and C2. (B) 31-mers that were significantly associated with CF and non-CF groups were aligned to the whole PAO1 genome and visualized by Phandango. Regions marked by asterisks are ribosomal DNA sequences and 'hs' marks the hotspot region where a large number of candidate 31-mers align. The higher the point representing a 31-mer, the more significant the Lrt p-value. (C) Proteins with large-scale mutation from the PAO1 genome and Lrt p-values calculated for clustering performed under two conditions (S95L95: 95% similarity and 95% length coverage; S70L95: 70% similarity and 95% length coverage) are described in the first and second columns as per the color index beneath the heatmap. 'hs' indicates candidate proteins located within the hotspot region in (B).

<https://doi.org/10.1371/journal.ppat.1009681.g001>

these studies are important in the context of adaptation to the CF lung. Here, we assembled a collection of 2,167 PA genomes from the Pseudomonas Genome Database [27], including genomes from both CF and non-CF individuals. Genomes of unknown origin were removed and a phylogeny was constructed using 1,001 genomes of known origin (Fig 1A). In order to remove clonal isolates that may increase any potential confounding lineage effects in the subsequent GWAS, the dataset was trimmed to 636 genomes (S1 Fig, S1 Data), maintaining 99.8% diversity of original selection and the overall structure of the phylogenetic tree. Isolates from CF patients could be identified from all parts of the phylogeny, including the previously described groups A and B; and the long branching groups C1 and C2 [28]. Most CF isolates (409 among 422) belonged to group A.

## Identification of CF-associated genetic elements

In total, 636 genomes were used to perform a GWAS using Pyseer [29]. These comprised 206 genomes from CF and 430 from non-CF. A k-mer based approach was implemented to include variability within intergenic regions. We selected k-mers with a length of 31 base pairs (31-mer) consistent with other studies that have adopted similar approaches and tested the effect of changes in k-mer length on GWAS sensitivity and false positive associations [22,23]. Significantly associated k-mers (p-value cutoff =  $2.8 \times 10^{-8}$ ) were aligned to the PAO1 reference genome and visualized using Phandango [30] (Fig 1B).

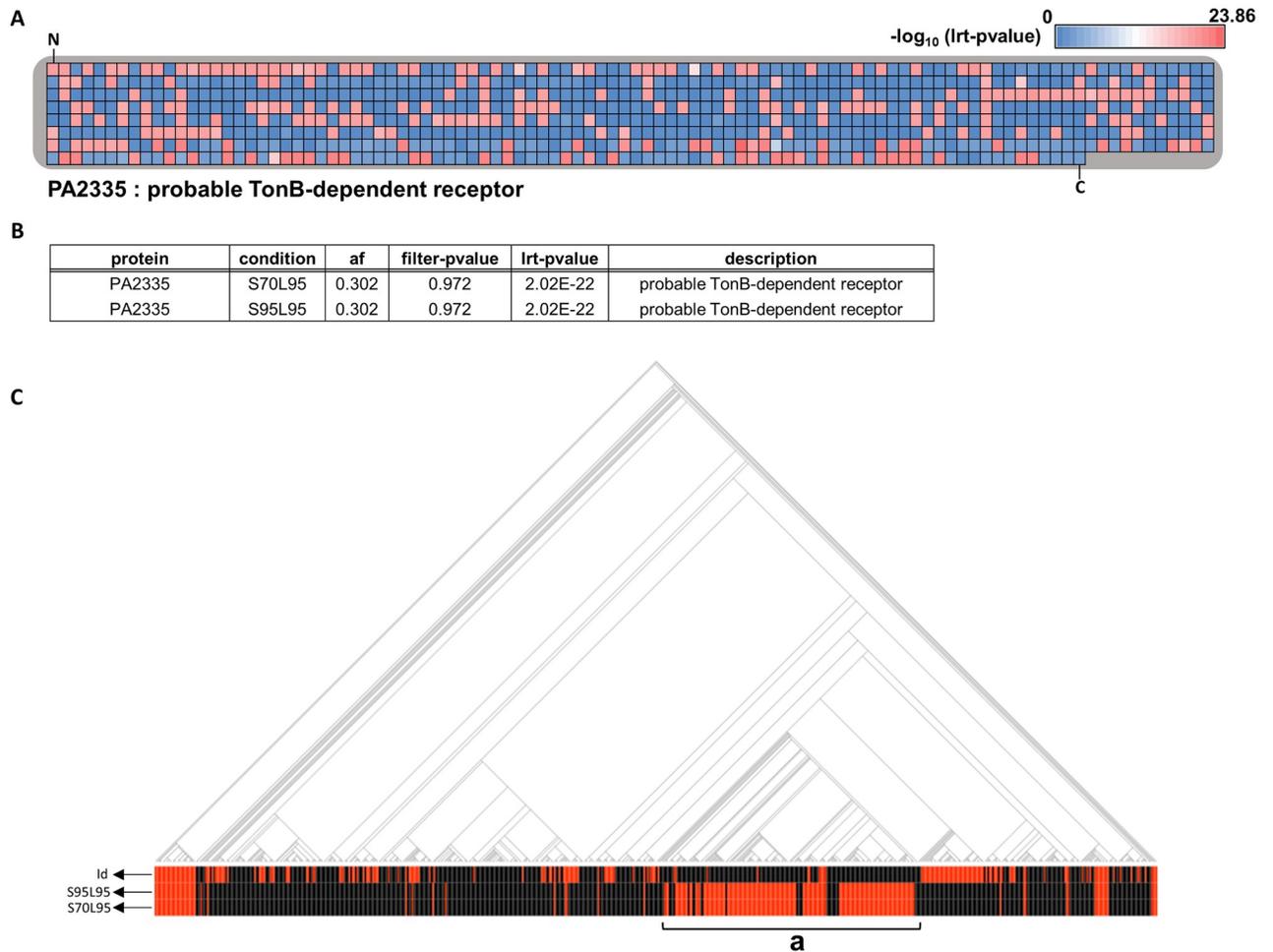
GWAS analysis identified 41,685 k-mers associated with CF with a p-value  $< 2.8 \times 10^{-8}$ . Of these, 489 *de-novo* assembled contigs mapped to 109 protein-coding genes, six untranslated regions and 29 intergenic regions in the PAO1 reference genome. Four most meaningful peaks marked with asterisks in Fig 1B are rDNA sequences. A greater number of CF isolates did not seem to have 16S and 23S rDNA sequences compared with the non-CF group. However, we verified that several of these CF isolates (AU2342, AU6462, AU9739, AU10409, AU16821 and AU25116) do have 16S rDNA by PCR-amplifying and sequencing the 16S rDNA regions. By extension, we believe that the lack of detection of 16S and 23S rDNA sequences from several CF isolates have arisen from an *in silico* error. Despite this, the number of genes annotated using Prokka [31] were not significantly different between genomes and they were therefore included in the input database.

## GWAS elements associated with allelic variation in *P. aeruginosa* from CF patients

Focusing on the k-mers that were mapped to the PAO1 reference genome, we obtained amino acid sequences for each gene from all isolate genomes. Two different homology clustering thresholds were used: 95% similarity and 95% gene length coverage (S95L95) and 70% similarity and 95% gene length coverage (S70L95) (S2 and S3 Data). GWAS results were then filtered by allele frequency using the likelihood-ratio test (lrt) p-values in Pyseer for both conditions (Fig 1C). This provided a list of candidate genes with multiple associated alleles. We hereinafter refer to these genes with multiple co-localized mutations to contain large-scale mutations. Variation in MucA (PA0763) served as a positive control, as MucA mutation is frequently observed in CF isolates [32,33].

## Large-scale alterations in protein-coding regions

Most candidate associated genes were identified within the prominent *hs* region (Fig 1C), with the most significant lrt p-value (p-value =  $2.02 \times 10^{-22}$ ) associated with a probable TonB-dependent receptor (S2 Data). Differences in the strength of association were noted across amino acid residues for PA2335 (Fig 2A). Analysis of raw filter p-values seemed to suggest an opposing conclusion in relation to the importance of PA2335 alteration in PA adaptation to the CF airway. However, these raw p-values are not weighted to account for the population structure (Fig 2B and 2C). Mapping variations in PA2335 against the phylogeny demonstrated alignment of the CF phenotype with the presence of its homolog, except for region a (Fig 2C). This region corresponds to the group B isolates (S1 Fig) and suggests that these isolates possess alternative CF associations. This pattern was also seen for variants in the PA2333, PA2334, and PA2336 genes. Variation in PA2332 was associated with the CF phenotype across the entire population suggesting it may have a broad influence in CF adaptation (S2 Data). This is consistent with a previous study that identified lineage-specific markers for groups A and B, including genes PA2333~PA2336 which were over-represented in group A [28].



**Fig 2. PA2335 candidate with the lowest lrt p-value in the large-scale mutation group.** (A) Total amino acid residues of PA2335 are presented and each amino acid locus is represented by individual squares. The color of each square indicates the lrt p-value assigned to that amino acid residue based on Pyseer results and colors are assigned according to the index above the heatmap, which was generated based on the negative logs of the maximum and minimum lrt p-values from the Pyseer result of amino acid locus. 'N' and 'C' mark the N- and C-terminus of PA2335, respectively. (B) Pyseer result of the PA2335 cluster is shown. Different conditions used in clustering and the frequency of the PA2335 large alteration are described in 'condition' and 'af' columns. Population structure is not considered in the calculation of filter p-values, but it is for the calculation of lrt p-values. (C) The phylogenetic tree of 636 genomes and their associated disease status are shown. Red and black lines in the 'id' row indicate genomes isolated from CF and non-CF patients, respectively. Black lines in the second and third rows indicate that the homolog of PA2335 clustered with the PAO1 protein at given clustering thresholds (S95L95:S70L95). Red lines indicate that the homolog did not cluster with the PAO1 protein. Genomes in region **a** are included in group B indicated in S1 Fig.

<https://doi.org/10.1371/journal.ppat.1009681.g002>

Functional annotation of proteins with large-scale mutations was performed using BlastKOALA [34]. KEGG ontology (KO) [35], definition, and pathway information of the candidate associated genes were identified in the CF or non-CF isolates (Table 1, S3 Data). Almost all of the large-scale mutations were more significantly associated with PA from CF patients. Large scale mutations of MucA (PA0763) are well known in CF isolates to result in increased alginate production and thus contribute to the mucoid phenotype, promoting infection [32,33,36,37]. PA2020 is a negative regulator of the MexXY efflux pump, which expels antibiotics from the cell [38]. Increased MexXY expression level and antibiotic resistance are induced by disruption to the PA2020 regulator. Moreover, CF sputum contains mucins, free amino acids, lactate, and glucose as a potential energy source and PA preferentially uses organic and amino acids over glucose as an energy source [39]. For this reason, we speculate that some CF

**Table 1. KEGG annotation of candidate proteins with large-scale mutation.** All large-scale mutations were annotated by Blastkoala. Proteins with KEGG information are listed in the 'Query' column, and KEGG ontology (KO), definition, and related pathway information are shown. Proteins discussed in the results are highlighted in bold.

Query	KO	Definition	Pathway
PA0763	<b>K03597</b>	<b>sigma-E factor negative regulatory protein RseA</b>	
PA1091	K20444	O-antigen biosynthesis protein	
PA2020	<b>K18129</b>	<b>TetR/AcrR family transcriptional regulator, mexXY operon repressor</b>	<b>Beta-Lactam resistance</b>
PA2320	K06145	LacI family transcriptional regulator, gluconate utilization system Gnt-I transcriptional repressor	
PA2321	<b>K00851</b>	<b>gluconokinase</b>	<b>Pentose phosphate pathway</b>
PA2322	K03299	gluconate:H <sup>+</sup> symporter, GntP family	
PA2323	K00131	glyceraldehyde-3-phosphate dehydrogenase (NADP <sup>+</sup> )	Glycolysis / Gluconeogenesis, Pentose phosphate pathway
PA2327	K02050	NitT/TauT family transport system permease protein	
PA2328	K02051	NitT/TauT family transport system substrate-binding protein	
PA2329	K02049	NitT/TauT family transport system ATP-binding protein	
PA2335	K02014	iron complex outer membrane receptor protein	
PA2338	<b>K10227</b>	<b>sorbitol/mannitol transport system substrate-binding protein</b>	<b>ABC transporters</b>
PA2340	<b>K10229</b>	<b>sorbitol/mannitol transport system permease protein</b>	<b>ABC transporters</b>
PA2341	<b>K10111</b>	<b>multiple sugar transport system ATP-binding protein</b>	<b>ABC transporters</b>
PA2342	<b>K00045</b>	<b>mannitol 2-dehydrogenase</b>	<b>Fructose and mannose metabolism</b>
PA2343	K00854	xylulokinase	
PA2344	K00847	fructokinase	Fructose and mannose metabolism
PA2345	K17218	sulfide:quinone oxidoreductase	Sulfur metabolism
PA2348	K20938	long-chain alkane monooxygenase	
PA2349	K02073	D-methionine transport system substrate-binding protein	ABC transporters
PA2350	K02071	D-methionine transport system ATP-binding protein	ABC transporters
PA2351	K02072	D-methionine transport system permease protein	ABC transporters
PA2352	K01126	glycerophosphoryl diester phosphodiesterase	Glycerophospholipid metabolism
PA3193	<b>K00845</b>	<b>glucokinase</b>	<b>Glycolysis / Gluconeogenesis</b>
PA3540	K00066	GDP-mannose 6-dehydrogenase	Fructose and mannose metabolism
PA3617	K03553	recombination protein RecA	Homologous recombination
PA3687	K01595	phosphoenolpyruvate carboxylase	Pyruvate metabolism
PA4216	K20262	dihydrophenazinedicarboxylate synthase	Phenazine biosynthesis, Quorum sensing
PA4701	K07028	uncharacterized protein	
PA4963	K09857	uncharacterized protein	
PA5025	K01740	O-acetylhomoserine (thiol)-lyase	Cysteine and methionine metabolism

<https://doi.org/10.1371/journal.ppat.1009681.t001>

isolates have evolved with mutations in glucose catabolism genes (PA2321 and PA3193) to avoid spending unnecessary energy in catabolizing a less optimal carbon source (glucose). Additionally, mannitol has been used as a therapeutic treatment to increase mucociliary clearance in the airway and simultaneously induce tobramycin sensitivity in PA by generating a proton motive force [40]. Therefore, we postulate that exposure through the concurrent therapeutic use of mannitol and antibiotics in CF patients may be the background from which the mutations arose in the gene cassette (PA2338, PA2340, PA2341, and PA2342). These large-scale mutations are highlighted in bold in [Table 1](#).

### Small-scale alterations in protein-coding regions

Large-scale mutations are more likely to cause loss-of-function than scattered substitutions in homologous sequence (small-scale mutation). However, adaptation to the CF lung may

**Table 2. Top 10 amino acid residues of the small-scale mutation group with 31-mers aligned to the PAO1 genome.** Top 10 amino acid residues according to Pyseer results with 31-mers aligned to the PAO1 genome are presented. The reference protein corresponding to each residue is presented in the ‘protein’ column, and the ‘locus’ column contains the locus of the mutation and amino acid residue of the reference protein where the mutation is detected. SLR sequences of PA5438 are highlighted in bold. Multiple types of mutation at each locus of the reference protein may be present and detailed information is provided in [S3 Data](#).

protein	locus	af	filter-pvalue	lrt-pvalue	description
<b>PA5438</b>	<b>272S</b>	<b>0.075</b>	<b>7.03E-20</b>	<b>9.92E-17</b>	<b>Probable transcriptional regulator</b>
<b>PA5438</b>	<b>273L</b>	<b>0.079</b>	<b>1.20E-19</b>	<b>4.41E-15</b>	<b>Probable transcriptional regulator</b>
<b>PA5438</b>	<b>274R</b>	<b>0.08</b>	<b>6.33E-19</b>	<b>3.36E-14</b>	<b>Probable transcriptional regulator</b>
PA4914	73A	0.05	2.71E-14	7.30E-13	Transcriptional regulator, AmaR
PA3355	384D	0.05	2.71E-14	3.48E-12	Hypothetical protein
PA2354	362G	0.09	3.36E-13	2.27E-11	SfnR1
PA2354	369E	0.08	2.20E-14	3.33E-11	SfnR1
PA2354	370R	0.08	2.20E-14	3.33E-11	SfnR1
PA2354	371A	0.08	2.20E-14	3.33E-11	SfnR1
PA2354	373R	0.08	2.20E-14	3.33E-11	SfnR1

<https://doi.org/10.1371/journal.ppat.1009681.t002>

also involve substitutions that alter protein-coding regions leading to additive effects. To investigate this, we analyzed the top 10 amino acid residues of the small-scale mutations (Tables 2 and 3). For the genes containing the most significant CF-associated variation, the functional genomics was further analyzed with laboratory validation experiments.

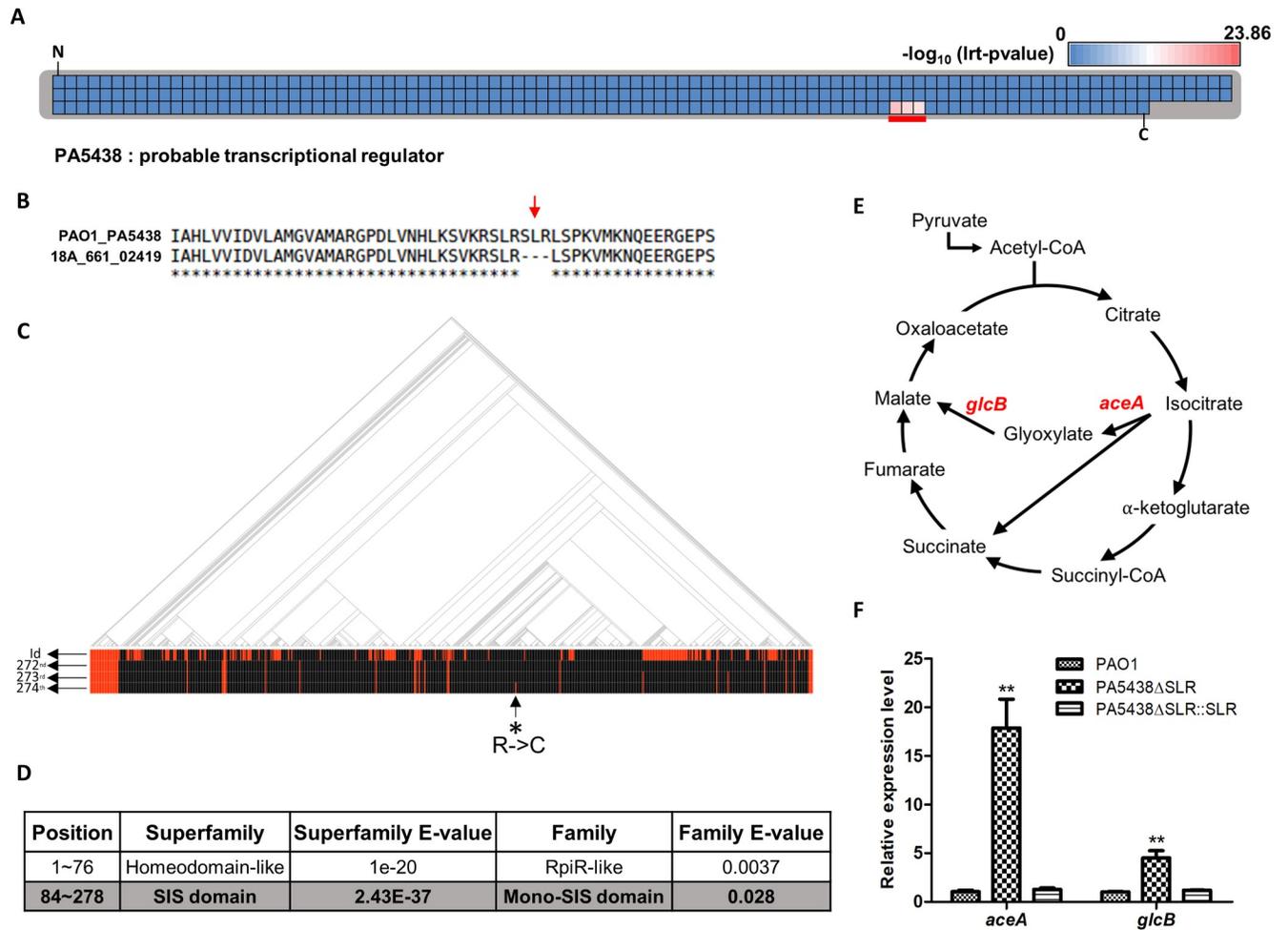
### Mutations in a putative transcriptional regulator (PA5438)

**Identification and experimental validation.** Mutations at amino acid loci 272–274 (SLR) in a probable transcriptional regulator gene (PA5438) were strongly associated with CF isolates (Table 2 and Fig 3A). The amino acid sequence ‘SLR’ was deleted in the 2,419<sup>th</sup> gene of the CF isolate 18A\_661, a PA5438 homolog (Fig 3B). The SLR deletion in the PA5438 protein exhibits a highly positive correlation with CF isolates overall (Fig 3C), and the PA5438 sequences (except for the SLR sequence region) were exactly identical to the reference PA5438 sequence. Domain search with SUPERFAMILY 2 [41] predicted these amino acids to be a part of a sugar isomerase (SIS) domain (Fig 3D). Therefore, we constructed an in-frame deletion mutant,

**Table 3. Top 10 amino acid residues of the small-scale mutation group whose 31-mers did not align to the PAO1 genome.** Top 10 amino acid residues according to Pyseer results whose 31-mers did not align to the PAO1 genome are shown. Reference proteins from genomes other than PAO1 and the PAO1 homolog of this reference protein are presented in columns ‘protein’ and ‘homolog’. The ‘locus’ indicates the mutation and amino acid of the reference protein where the mutation is detected. Multiple types of mutation at each locus of the reference protein may be present and detailed information is provided in [S3 Data](#). The reference protein name in the “protein” column is composed of the genome id, connected by the latter underscore sign, to gene number within that genome. All sequences of the reference proteins in Table 3 are provided in [S4 Data](#).

Protein	locus	af	filter-pvalue	lrt-pvalue	homolog (PAO1)	description
<b>AU17965_3981_04951</b>	<b>162S</b>	<b>0.0708</b>	<b>1.27E-19</b>	<b>3.80E-16</b>	<b>PA0313</b>	<b>L-cystine transporter of ABC system YecS</b>
<b>AU17965_3981_04951</b>	<b>163L</b>	<b>0.0708</b>	<b>1.27E-19</b>	<b>3.80E-16</b>	<b>PA0313</b>	<b>L-cystine transporter of ABC system YecS</b>
<b>AU17965_3981_04951</b>	<b>164I</b>	<b>0.0708</b>	<b>1.27E-19</b>	<b>3.80E-16</b>	<b>PA0313</b>	<b>L-cystine transporter of ABC system YecS</b>
105738_3985_01941	57G	0.95	2.71E-14	1.12E-12	PA1384	UDP-glucose 4-epimerase
AZPAE14712_2411_01394	206G	0.928	2.19E-16	1.38E-09	PA2353	Conserved hypothetical protein
AZPAE14712_2411_01394	198P	0.0739	1.01E-15	4.72E-09	PA2353	Conserved hypothetical protein
AZPAE14712_2411_01394	200R	0.0739	1.01E-15	4.72E-09	PA2353	Conserved hypothetical protein
AZPAE15072_2259_00589	283T	0.0503	2.71E-14	1.70E-08	PA3848	Hypothetical protein
AZPAE15072_2259_00888	176D	0.0503	2.71E-14	1.70E-08	PA3594	Probable transcriptional regulator
AZPAE15072_2259_02084	305K	0.0503	2.71E-14	1.70E-08	PA0242	Hypothetical protein

<https://doi.org/10.1371/journal.ppat.1009681.t003>

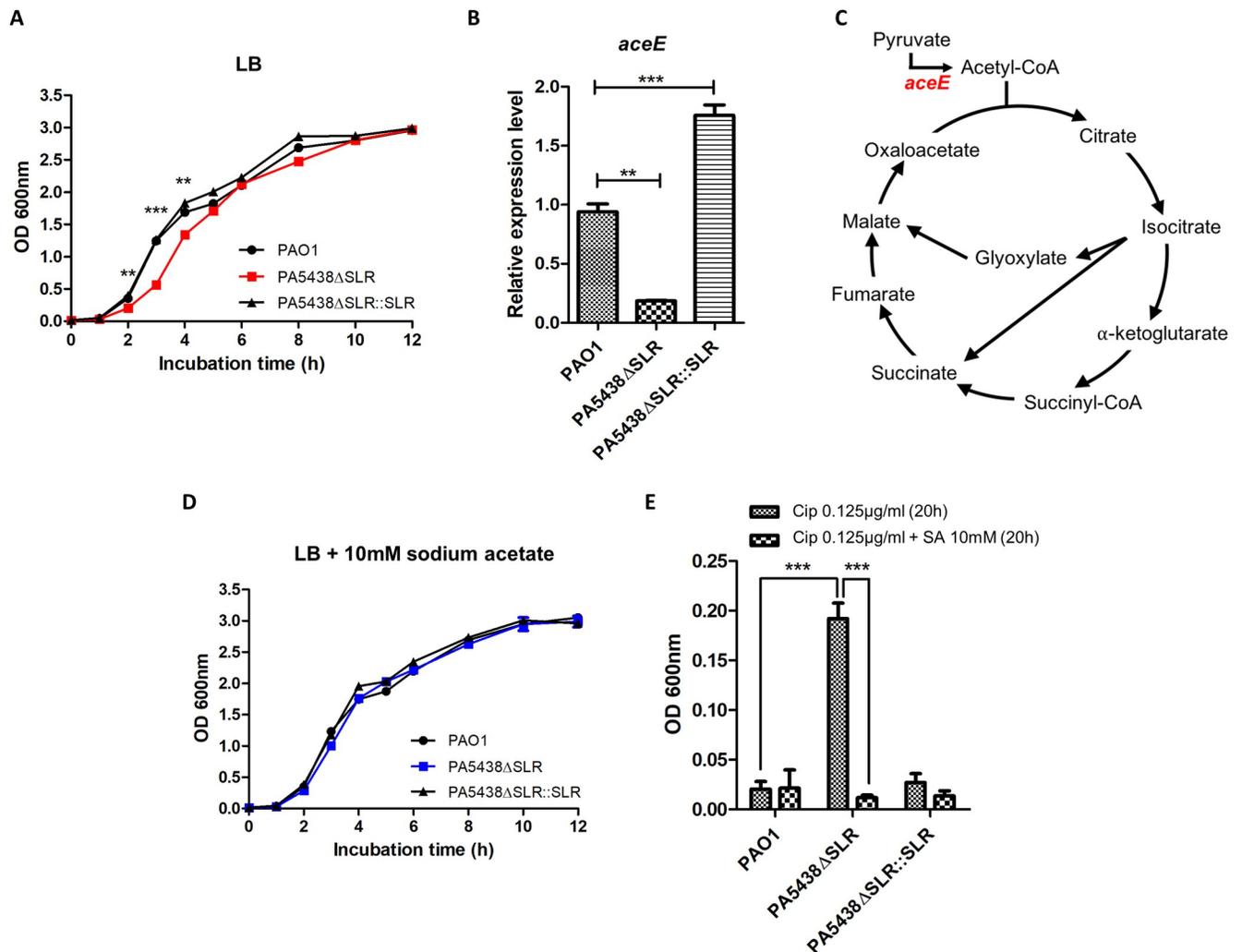


**Fig 3. SLR deletion in PA5438 increases the expression level of *aceA* and *glcB*.** (A) Total amino acid residues of PA5438 are presented and each amino acid locus is represented by individual squares. The color of each square indicates the Irt p-value assigned to that amino acid residue based on Pyseer results and colors are assigned according to the index above the heatmap, which was based on the negative logs of the maximum and minimum Irt p-values from the Pyseer result of amino acid locus. 'N' and 'C' mark the N- and C-terminus of PA5438, respectively. Residues at locus 272–274 of PA5438 with significant Irt p-values are underlined in red. (B) Comparison of the reference PA5438 to a homolog protein from a CF-isolated genome (18A\_661\_02419) is shown. The red arrow indicates deletion of SLR residues in PA5438 and corresponds to the red-underlined locus observed in A. (C) Phylogenetic tree of 636 genomes and their associated disease status are shown. Red and black lines in the 'id' row indicate genomes isolated from CF and non-CF patients, respectively. Red lines in the second, third, and fourth rows (272<sup>nd</sup>, 273<sup>rd</sup>, and 274<sup>th</sup>) each correspond to the absence of S, L, R residues, respectively. One exception to this is the red line in row 274 (asterisk-arrow), which was detected as R substituted by C. (D) A domain search result of PA5438 by SUPERFAMILY 2 is shown and the position of the SLR sequence inside the SIS domain is highlighted in bold. (E) AceA (isocitrate lyase) and GlcB (malate synthase) are enzymes involved in the glyoxylate shunt pathway (highlighted in red). (F) RNAs of PAO1, the PA5438ΔSLR mutant, and the complementation strain were extracted at OD<sub>600nm</sub> ~1.0 in LB medium and relative expression levels of *aceA* and *glcB* were measured. \*\**p*<0.01.

<https://doi.org/10.1371/journal.ppat.1009681.g003>

PA5438ΔSLR, to determine whether these three amino acids affect the transcriptional regulator function or not.

PA5438 is a transcriptional repressor that directly binds to the promoter region of the *aceA* (isocitrate lyase) gene and has been shown to repress the expression of the *glcB* (malate synthase) gene during growth in a non-C<sub>2</sub> carbon source [42]. These repressed genes encoding enzymes involved in the glyoxylate shunt (GS) pathway (Fig 3E). To determine whether the suppressive activity of PA5438 is lost in the PA5438ΔSLR mutant, gene expression levels of *aceA* and *glcB* were measured by quantitative real-time PCR (qRT-PCR). The qRT-PCR results show that the expression level increased 17-fold for *aceA* and 5-fold for *glcB* (Fig 3F). Based on



**Fig 4. SLR deletion in PA5438 affects antibiotic susceptibility in LB medium.** (A) Growth curves of PAO1, PA5438ΔSLR mutant, and the complementation strain in LB were observed over 12 hours.  $**p < 0.01$ ;  $***p < 0.001$  (B) RNAs of PAO1, PA5438ΔSLR mutant, and the complementation strain were extracted at  $OD_{600nm} \sim 1.0$  in LB and relative expression levels of *aceE* were measured.  $**p < 0.01$ ;  $***p < 0.001$  (C) AceE (pyruvate dehydrogenase) is an enzyme involved in the entrance of The Krebs Cycle (highlighted in red). (D) Growth curves of PAO1, PA5438ΔSLR mutant, and the complementation strain in LB supplemented with 10mM sodium acetate were observed over 12 hours. (E) Antibiotic susceptibility tests against ciprofloxacin (Cip) and 10 mM sodium acetate (SA) were performed. Initial  $OD_{600nm}$  of PAO1, mutant, and the complementation strain were adjusted to 0.005, and  $OD_{600nm}$  was measured after 20 hours of static incubation in LB supplemented with Cip. The concentration of antibiotic was 0.125 μg/ml.  $***p < 0.001$ .

<https://doi.org/10.1371/journal.ppat.1009681.g004>

these findings, we expect that the deleted SLR sequence in PA5438 is an important region for determining *aceA* and *glcB* expression levels.

Further investigation of the PA5438ΔSLR mutant was performed to identify phenotypes that may aid in the adaptation of the strain to the CF environment. First, we assessed the phenotype of the mutant PA5438ΔSLR in LB media as growth rates of diverse CF isolates in LB media closely approximate those of CF isolates grown in artificial sputum medium (ASM) and synthetic CF sputum medium (SCFM) [17]. Slower growth in LB was observed in the PA5438ΔSLR mutant compared to PAO1 and the complementation strain, especially during the exponential phase (Fig 4A). In a recent study, RccR of *Pseudomonas fluorescens*, a PA5438 homolog, was shown to control pyruvate metabolism as well as the GS pathway [43]. Interestingly, pyruvate dehydrogenase (*aceE*) and *aceA* were regulated differently in the presence of

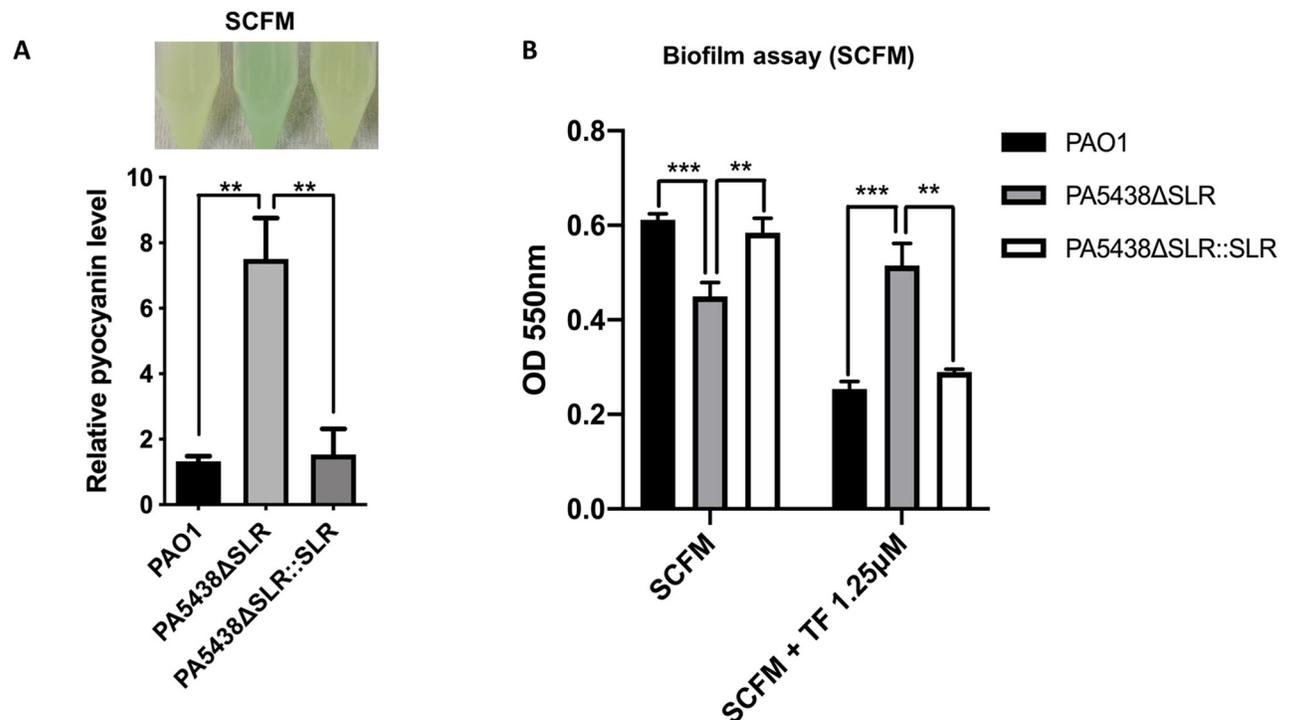
2-keto-3-deoxy-6-phosphogluconate (KDPG), which is bound to RccR [43]. We measured the relative expression level of *aceE* at the exponential phase because pyruvate metabolism affects growth. The expression of *aceE* in the mutant was suppressed compared to wildtype PAO1 and the complementation strain (Fig 4B). AceE generates acetyl-CoA from pyruvate, which feeds into the TCA cycle (Fig 4C). Therefore, we considered if the addition of metabolites related to this enzyme reaction (either as a precursor or product) would achieve recovery of growth. A defect in mutant growth was still observed with the addition of sodium pyruvate and sodium citrate, but the addition of sodium acetate as a substitute of acetyl-CoA resulted in a complete recovery of growth (Fig 4D).

Since antibiotic tolerance can be associated with slow-growing or non-dividing “persister” bacteria [44], we measured susceptibilities of PAO1, PA5438ΔSLR and the complementation strain to different classes of antibiotics (tobramycin and ciprofloxacin) commonly used to treat PA infection in CF patients. In static cultures with tobramycin or ciprofloxacin, the PA5438ΔSLR mutant reached a significantly higher OD<sub>600nm</sub> in ciprofloxacin compared to wildtype PAO1 and the complementation strain, but only slightly higher OD<sub>600nm</sub> in tobramycin (Figs 4E and S2A). Since growth in LB recovered with sodium acetate (Fig 4D), we performed antibiotic susceptibility tests with sodium acetate supplementation. A limited increase of tobramycin susceptibility was observed in PA5438ΔSLR (S2A Fig), whereas susceptibility to ciprofloxacin was fully restored to that of the wildtype PAO1 and complementation strain (Fig 4E). Therefore, we postulate that slower growth is involved in the decreased susceptibility of the PA5438ΔSLR mutant to the two antibiotics.

Evidently, the composition of nutrients in LB media is very different to that of the CF sputum. Therefore, a synthetic medium was prepared to mimic the nutritional makeup of CF sputum (SCFM) and used it to measure growth. Unlike growth in the LB medium, in SCFM, wildtype PAO1, PA5438ΔSLR and the complementation strain showed similar growth (S2B Fig). Since the decreased expression level of *aceE* seemed to be an important factor of growth in LB medium, we measured the expression levels of *aceE* in PA cultured in SCFM. The expression level of *aceE* in the PA5438ΔSLR mutant was similar to that of the wildtype PAO1 and complementation strain (S2C Fig). In contrast, *aceA* and *glcB* expression levels were significantly increased in the PA5438ΔSLR mutant compared to the wildtype and complementation strain grown in SCFM (S2D Fig).

PA produces pyocyanin, an important virulence factor derived from phenazine and forms blue-green pigments. Pyocyanin induces reactive oxygen species (ROS) generation by transferring electrons to oxygen and increasing neutrophil apoptosis as a way of disrupting the host immune system [45]. As such, the supernatant color of a PA culture can function as a proxy for the virulent nature of the bacterium. Interestingly, we observed the 8 hour culture supernatant of PA5438ΔSLR in SCFM to be greener than that of the PAO1 and the complementation strain (Fig 5A). When we confirmed the relative pyocyanin levels of this bacterial culture, the PA5438ΔSLR mutant in SCFM resulted in 6~7 times more pyocyanin than PAO1 and the complementation strain (Fig 5A). Pyocyanin stimulates iron removal from transferrin in bacteria incubated under low oxygen conditions [46], and the level of transferrin is increased in the bronchoalveolar fluid from CF patients, compared to that of healthy individuals [47]. Therefore, we measured biofilms in which oxygen is limited, using the iron-saturated holo-transferrin. Compared to PAO1 and the complementation strain, the PA5438ΔSLR mutant produced less biofilm in SCFM, as demonstrated by a decreased OD<sub>550nm</sub> absorbance value. However, in the presence of transferrin, the mutant produced a similar level of biofilm whereas the PAO1 and the complementation strain produced less biofilms than in SCFM (Fig 5B).

**Potential implications in CF airway infections.** These findings have potential implications in CF airway infections. An altered fatty acid profile has been associated with the airways



**Fig 5. SLR deletion in PA5438 affects virulence and biofilm production in SCFM.** (A) After bacterial culture of PAO1, PA5438ΔSLR mutant, and the complementation strain in SCFM for 8 hours with shaking, a pyocyanin assay was performed with culture supernatant, and relative pyocyanin level was calculated using PAO1 pyocyanin level. Bacterial cultures used in pyocyanin assay are shown.  $**p < 0.01$  (B) After 32 hours of static culture of PAO1, PA5438ΔSLR mutant and the complementation strain in SCFM and SCFM supplemented with holo-transferrin (TF), biofilm was measured spectrophotometrically.  $**p < 0.01$ ;  $***p < 0.001$ .

<https://doi.org/10.1371/journal.ppat.1009681.g005>

of CF patients. For example, higher concentrations of palmitic acid and oleic acid have been detected in CF airway samples [48]. Fatty acid degradation upregulates the glyoxylate shunt (GS) pathway [49], and PA isolated from the CF lungs can induce the expression of genes involved in fatty acid metabolism and the GS pathway [50]. Furthermore, mucin is a major energy source in the CF lung environment in addition to fatty acids, and the GS pathway is important for mucin degradation and consumption [51]. Meta-transcriptomic analysis of several CF sputum samples in another study also found that *aceA* was upregulated and genes associated with glucose transporters and glycolysis were significantly downregulated [52]. Based on these previous studies, we assumed that the activation of the GS pathway may provide an advantage for PA proliferation in CF lungs.

Previous observations of GS pathway upregulation [50,52] are probably due to the presence of CF isolates that contain the abnormal PA5438 gene. In the PA5438ΔSLR mutant, *aceA* and *glcB* expression levels were increased not only in LB but also in SCFM. To investigate if the end products of the upregulated expression of *aceA* and *glcB* are functional, we examined the sequences of AceA and GlcB from genomes that harbor ΔSLR. Both promoters and proteins of AceA and GlcB were highly conserved (Table A, Table B in S1 Table). This indicates that the SLR deletion is the sole contributor to the transcriptional regulation of these two genes.

Glyoxylate shunt pathway is necessary for surviving phagocytosis and is upregulated in response to ROS [53,54]. We performed a gentamicin protection assay using our strains and mouse bone-marrow-derived neutrophils to measure how well the PA5438 mutant endures ROS-enriched and nutrient-poor phagocytic environment. We did not observe a significant

difference in the CFUs of the PA5438 mutant strain compared to the wild type and the complementation strains in our experiment (S2E Fig).

As CF isolates have similar growth rates in LB and SCFM [17], we chose LB as the medium in which to investigate phenotypic differences between the wildtype and mutant PA. However, we observed a marked difference between the growth of the PA5438 $\Delta$ SLR mutant cultured in LB and SCFM. In addition, the bacterial culture had a greener colour when SCFM was used (Fig 5A) than when LB was used. Given that PA5438 is involved in energy metabolism [43], we hypothesized that the mutant and wildtype phenotypes may vary depending on the type of medium used (LB or SCFM). Therefore, virulence phenotypes of the mutant, such as the increased pyocyanin levels observed in SCFM, are more likely to occur in the CF environment than the decreased antibiotic susceptibility observed in LB cultures. We also examined the genes involved in pyocyanin biosynthesis in the genomes that carry the PA5438 $\Delta$ SLR mutation. Whereas several genomes were found to carry mutations in several of these genes when compared with the PAO1 sequences, other genomes possess a highly conserved gene set (S3 Fig). Based on this finding, we anticipate that virulent phenotypes such as that of the PA5438 $\Delta$ SLR mutant are likely to manifest in the CF lung environment.

Despite differences in the nutrient compositions of LB and SCFM, the mutant phenotypes observed in LB cultures provide evidence that PA5438 acts as a transcriptional regulator. RccR of *Pseudomonas fluorescens*, a PA5438 homolog, has been shown to possess an SIS domain, the binding affinity of which is regulated by 2-keto-3-deoxy-6-phosphogluconate (KDPG) [43]. This KDPG-dependent moderation of PA5438 binding affinity is also dependent on the gene with which PA5438 binds. In the absence of KDPG, the binding affinity of PA5438 to *aceE* is increased, whereas its binding affinity to *aceA* is decreased [43]. A domain search using the SUPERFAMILY 2 database detected the SLR amino acid sequence of PA5438 to be inside the SIS domain (Fig 3D). The qRT-PCR results showed a decreased *aceE* gene expression level and an increased *aceA* gene expression level in LB (Figs 3F and 4B). Based on these findings, we propose that the SLR amino acid sequence may be a key region to which KDPG binds. In contrast to LB culture, in SCFM culture the *aceE* gene expression by the PA5438 $\Delta$ SLR mutant is similar to that of wildtype PAO1 and the complementation strain. This observation may be attributed to the preferential use of amino acids over glucose in SCFM, which may lead to a decreased intracellular level of KDPG in the exponential phase of PA cultured in SCFM, unlike in LB. Further investigations are needed to conclusively determine whether KDPG indeed binds to the SLR region of PA5438.

Furthermore, elastase and pyocyanin are well-characterized virulence factors regulated by quorum sensing. We performed an elastase assay to test whether quorum sensing is related to the increased pyocyanin level [20]. Similar to the pyocyanin assay results, which showed a significant increase in pyocyanin in the PA5438 $\Delta$ SLR mutant, relative elastase activity was slightly increased in the mutant compared to wildtype PAO1 and the complementation strain (S2F Fig). Therefore, we anticipate the increased pyocyanin level to be mediated by the quorum-sensing system. Pyocyanin stimulates iron removal from transferrin when PA is grown in low oxygen conditions [46], and increased levels of transferrin have been observed in CF patients [47]. Biofilm is a matrix of extracellular polymeric substances which protects the bacterial cells beneath, and oxygen supply is limited in developed biofilms [55]. Biofilm production of the PA5438 $\Delta$ SLR mutant is decreased compared to PAO1 and the complementation strain in SCFM. In SCFM with transferrin, the mutant maintains the level of biofilm production, whereas PAO1 and the complementation strain produced less biofilms (Fig 5B). This implies that the mutant maintenance of biofilm production level in the presence of transferrin may confer greater fitness in the transferrin-enriched CF environment. Based on our experimental

results, we speculate that this mutant may be problematic in the CF lung due to its increased virulence and biofilms.

### Mutations in the L-cystine transporter (YecS, PA0313)

**Identification and experimental validation.** A major limitation of aligning 31-mers to the PAO1 reference genome is that the 31mers that do not align due to insertion mutations are overlooked. To determine the reference genes for such insertion mutations, un-aligned 31-mers were *de novo* assembled, and the contigs that formed were mapped to 635 genomes (excluding PAO1). The top 10 amino acid residues with corresponding 31-mers that do not align to the PAO1 genome are described in [Table 3](#). Three amino acid residues at loci 162–164 ([Fig 6](#)) had significant lrt p-values compared to other amino acid residues. At these loci, the YecS homolog had insertions of three amino acid residues, SLI ([Fig 6](#)). Moreover, there was an additional copy of SLI in the AU17965\_3981\_04951 protein, comprising a total of three stretches of ‘SLI’, in contrast to a total of two stretches in the PAO1 YecS ([S4 Fig](#)). SLI insertions were strongly associated with CF isolates ([Fig 6C](#)) and the YecS homologs of these isolates were identical to the reference AU17965\_3981 sequence, except for the additional SLI in PAO1 YecS. The AU17965\_3981\_04951 protein is a cytoplasmic-membrane transporter, and the additional ‘SLI’ sequence is thought to span both cytoplasmic and transmembrane regions ([S5A and S5B Figs](#)) [[56](#)].

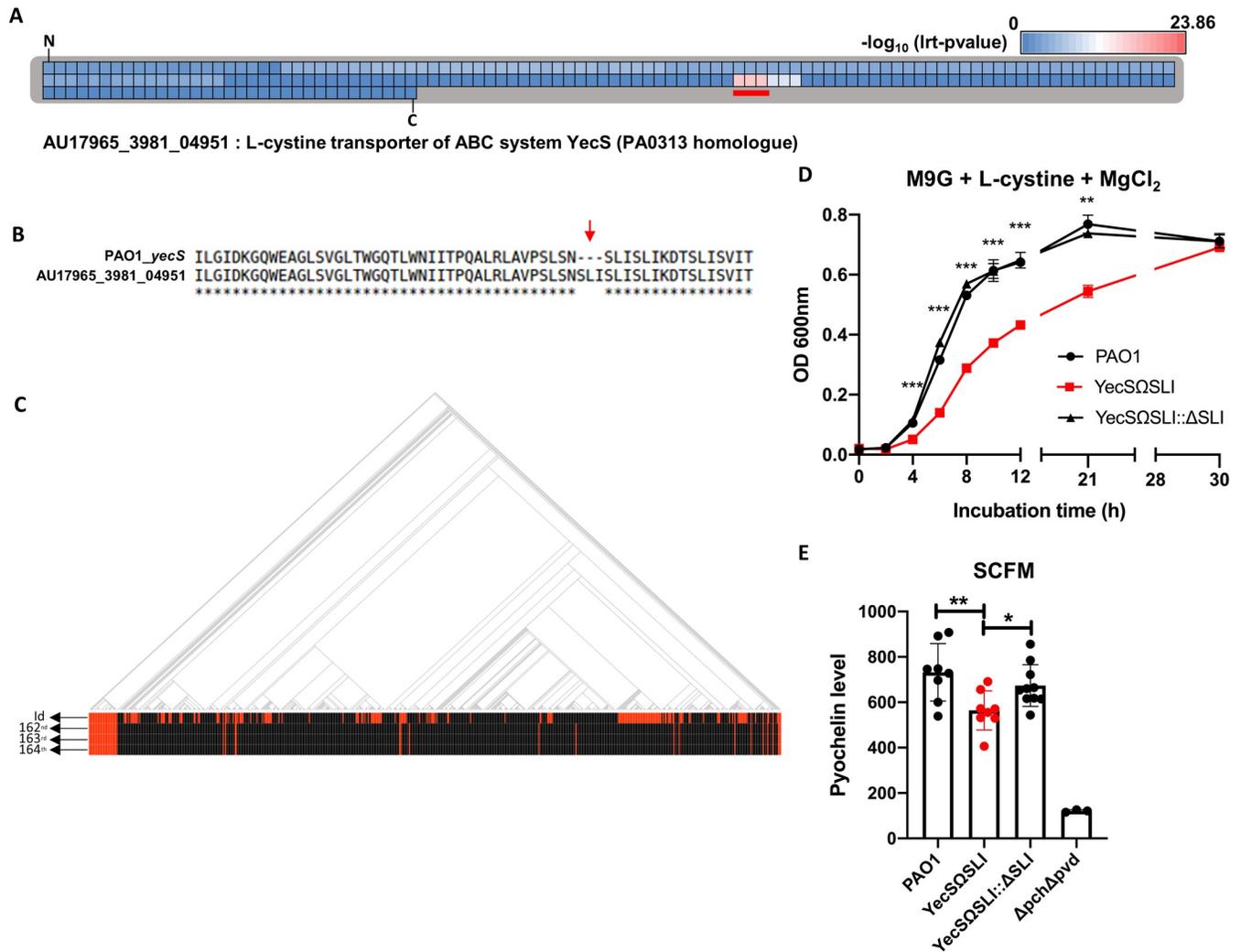
To experimentally assess if the SLI insertion effects a phenotypic change, an in-frame mutant was constructed with the additional SLI sequence inserted into the corresponding locus of PAO1 YecS. No difference in growth was observed between the wildtype PAO1 and the YecS $\Omega$ SLI mutant when both strains were grown in SCFM ([S5C Fig](#)). However, in M9 minimal media supplemented with glucose and L-cystine as the sole sulfur source, the mutant grew slower than PAO1 and the complementation strain, whereas growth in M9 media with an additional sulfur source present (MgSO<sub>4</sub>) showed no difference ([Figs 6D and S5D](#)).

Two L-cystine transporters are known in *E. coli*, an ATP-binding cassette (ABC) importer (FliY-YecSC) and symporter YdjN [[57](#)]. Previous work has shown that the presence of either of the two transporters is sufficient to meet the L-cystine requirement for normal growth when L-cystine is used as the sole sulfur source [[57](#)]. We performed a blastp search of 636 genomes using the database of 238 YdjN homologs [[58](#)] and concluded that 635 out of 636 genomes do not seem to encode a YdjN homolog ([S5 Data](#)). Therefore, based on our blast search and experimental result ([Fig 6D](#)), we hypothesize that PAO1 only possesses the ABC importer system for L-cystine uptake. Moreover, the decreased growth of the SLI insertion mutant is attributed to the decreased activity of the ABC transporter rather than a complete loss of function. A comparison of growth between a YecS clean deletion mutant and our SLI insertion mutant would help in better understanding the assortment of L-cystine transporters present in PA.

### Potential implications in CF airway infections

In addition to the SLI insertion mutation, we focused on other patterns of mutation in YecS that were observed across the CF isolate genomes. One such pattern is the deletion of an SLI sequence that results in CF isolates encoding just one SLI sequence in YecS homologs ([S4A Fig](#)). Another pattern of mutation observed by multiple alignments of YecS homologs is the deletion of long stretches of amino acid residues ([S4B Fig](#)). Since the SLI insertion mutant exhibited decreased L-cystine transporter activity ([Fig 6D](#)), we anticipate that such large deletions in the YecS protein would incur a complete loss of function.

Amino acid concentrations in the sputum vary across individual CF patients [[59](#)] and cysteine levels can be low or even undetectable [[59](#)]. Since cysteine is oxidized to cystine, a low level



**Fig 6. SLI insertion in YecS reduces L-cystine utilization activity.** (A) The amino acid residues of CF-derived AU17965\_3981\_04951 protein are presented and each amino acid locus is represented by individual squares. The color of each square indicates the lrt p-value assigned to that amino acid residue based on Pyseer results and colors were assigned according to the index above the heatmap. The color index was generated based on the negative logs of the maximum and minimum lrt p-values from the Pyseer result of the amino acid locus. 'N' and 'C' mark the N- and C-terminus of AU17965\_3981\_04951, respectively. Residues presenting especially low lrt p-values are marked with a red underline and were found at loci 162–164 of the AU17965\_3981\_04951. (B) Comparison of the reference AU17965\_3981\_04951 to its PAO1 homolog YecS (PA0313) is shown. The red arrow indicates insertion of SLI residues in AU17965\_3981\_04951 and corresponds to the red-underlined locus observed in (A). (C) Phylogenetic tree of 636 genomes and their associated disease status are shown. Red and black lines in the 'id' row indicate genomes isolated from CF and non-CF patients, respectively. Red lines in the second, third, and fourth rows indicate the presence of SLI insertion in PAO1 YecS. (D) Growth curves of PAO1, YecS $\Delta$ SLI mutant and the complementation strain cultured in M9 minimal media with MgCl<sub>2</sub> instead of MgSO<sub>4</sub>, supplemented with glucose and L-cystine as the sole sulfur source were measured over 30 hours. \*\* $p < 0.01$ ; \*\*\* $p < 0.001$  (E) After bacterial culture of PAO1, YecS $\Delta$ SLI mutant, the complementation strain and the  $\Delta$ pch $\Delta$ pvd mutant, a pyochelin- and pyoverdine- deficient strain, in SCFM for 6 hours with shaking, the pyochelin in the bacterial culture supernatant was measured. \* $p < 0.05$ ; \*\* $p < 0.01$ .

<https://doi.org/10.1371/journal.ppat.1009681.g006>

of one may indicate a low level of the other in the CF lung. Therefore, on one hand, large deletions in YecS may provide a survival advantage by allowing the CF isolate to avoid the energy cost of maintaining a functional YecS transporter in the cysteine-deficient CF lung environment. On the other hand, the reduction of intracellular cystine to cysteine contributes to ROS production in *E. coli* [60]. Thus, we expect small-scale mutations in YecS may be a result of evolution that protects the CF isolates from ROS stress in the cysteine-present CF lung environment. Additionally, pyochelin is a well-known siderophore produced by PA. Pyochelin

may inflict tissue damage in the CF lung by inducing a continuous inflammatory response [61]. Interestingly, cysteine is a precursor molecule required for the production of pyochelin [62]. Therefore, we tested the effect of SLI insertion in *YecS* on pyochelin production of PAO1, the *YecS*SLI mutant, the complementation strain, and a  $\Delta pch\Delta pvd$  mutant [63] that does not produce pyochelin and pyoverdine. The SLI insertion mutant exhibited decreased pyochelin production compared to the wildtype and the complementation strain (Fig 6E). A decreased level of pyochelin may be associated with a less virulent phenotype of the mutant, which may be an adaptation strategy for establishing a chronic infection in the CF lung environment. Whereas several genomes carried mutations in pyochelin biosynthesis genes, when compared with the PAO1 sequences, other genomes possess a highly conserved gene set (S6 Fig). Therefore, it may be the case that less virulent phenotypes such as that of the *YecS*SLI mutant proliferate in the CF lung environment.

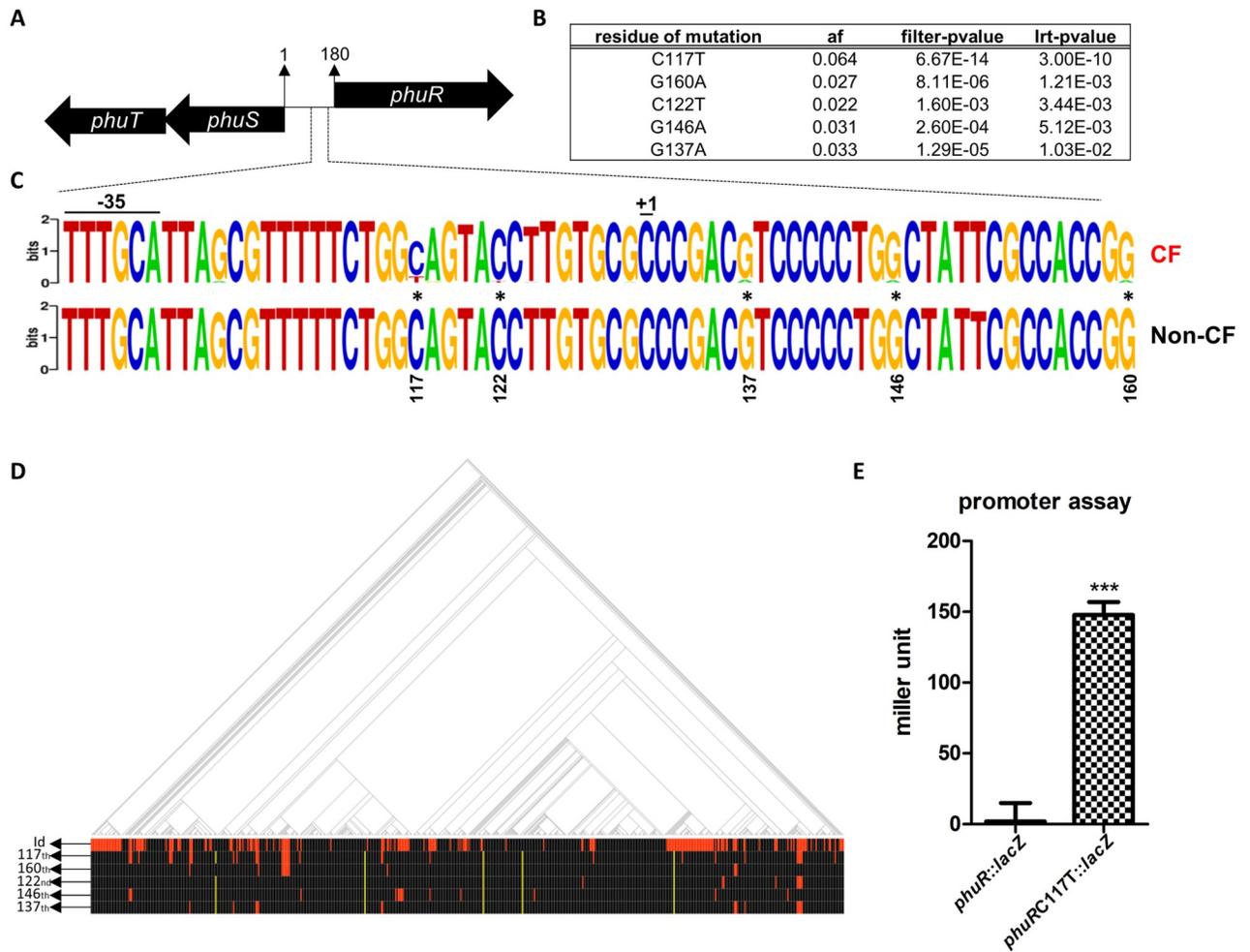
### Small-scale alterations in intergenic regions: *phuS* and *phuR*

Comparison of genomes using k-mers enables the investigation of intergenic regions. Of all contigs assembled from significantly different 31-mers, between CF and non-CF isolates, 29 contigs aligned to the non-coding regions of the PAO1 genome. Most aligned to either the *hs* region or rDNA sequences (Fig 1B). Since mutations located in the *hs* region may cause gene deletions, and rDNA sequence regions were excluded from analysis due to difficulties in interpretation, we selected regions other than these loci. Among several such intergenic regions, one intergenic region between *phuR* and *phuS* operons, was involved in the *pseudomonas* heme utilization (*phu*) systems (Fig 7A). The function of such a system is the acquisition of iron from the heme group of hemoglobin [64]. To evaluate the potential role of these mutations in CF adaptation, Pyseer analysis was performed using the nucleotide sequences of this intergenic region from PAO1 and other isolates (Fig 7B). The most significant hit was the transition of the 117<sup>th</sup> residue on the forward strand from cytosine (C) to thymine (T) (Fig 7B). Two mutations (C117T and C122T) are included in the *phuR* promoter region (from -35 region to transcriptional initiation site (+1)) (Fig 7C).

Consistent with our findings, the frequency of mutations within this intergenic region in a CF isolate were found to increase significantly compared to the background mutation rate, and the mutated intergenic region increased *phuR* promoter activity [65]. In some cases, the intergenic region was deleted (Fig 7D) and the promoter activity is probably absent in these isolates. When such cases of promoter deletion are excluded from the analysis, a stronger correlation is expected between the mutations of this region and CF isolates. To examine whether the C117T point mutation affects the expression of *phuR*, we performed a promoter activity assay by measuring  $\beta$ -galactosidase activity. Promoters with this mutation exhibited increased activity (Fig 7E). Therefore, increased *phuR* promoter activity will ultimately result in increased *phu* system activity. We speculate that the C117T mutation is an important strategy for iron uptake and thus survival in the CF environment.

### Conclusions

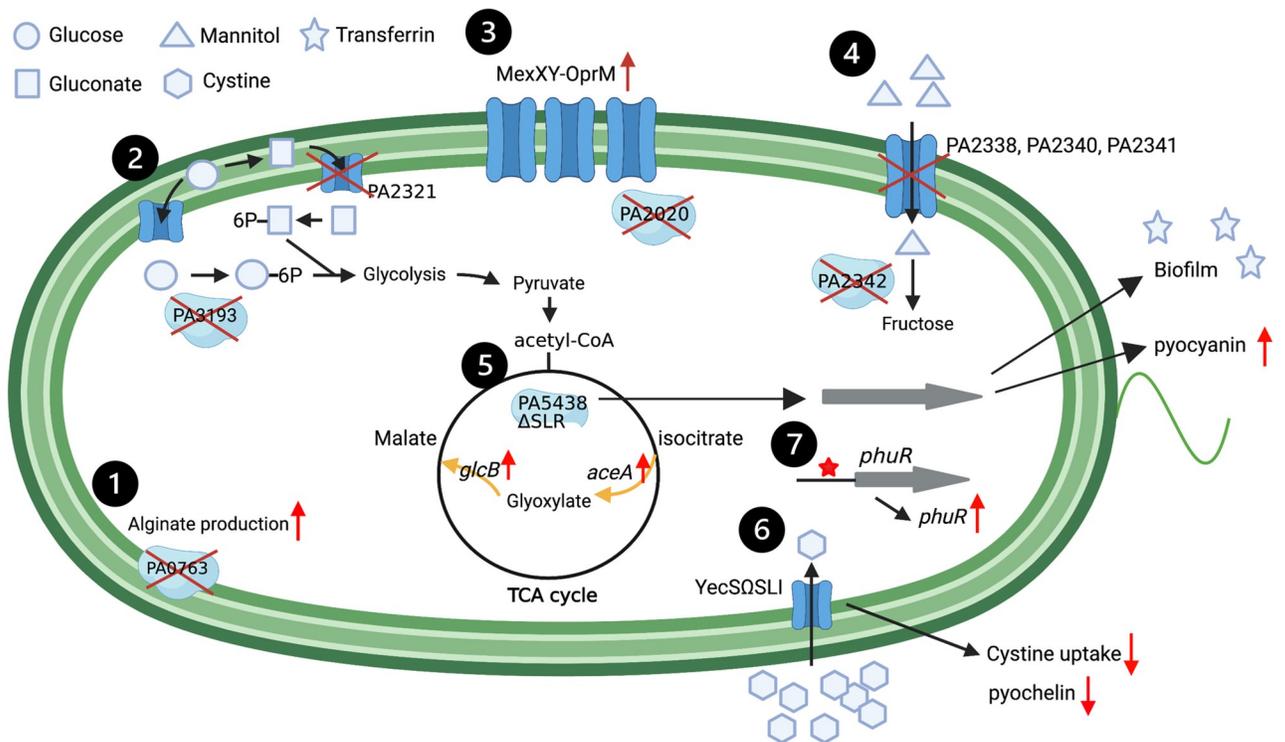
In this study, we compared a large set of genomes from clinical PA isolates (CF vs. non-CF) and identified mutations that occurred differentially in either CF or non-CF isolates using GWAS based on 31-mer counting (Fig 8). To integrate these findings within a functional genomics context [66], the impact on bacterial growth- and virulence-associated phenotypes was also assessed. Importantly, PAO1-derived variants, PA5438 $\Delta$ SLR and *YecS*SLI, exhibited distinct phenotypes in *in vitro* assays. Moreover, a single nucleotide replacement in the promoter region of the *phuR* gene caused a considerable increase in gene transcription. Together, these



**Fig 7. One SNP in the *phuR* promoter region increases expression level.** (A) The 180 bp intergenic region between *phuR* and *phuS* is shown. (B) The top 5 associated SNPs detected using Pyseer within this intergenic region are listed and the number in the ‘residue of mutation’ column indicates the forward strand-based locus of mutation. The nucleotide in front of the locus number indicates the reference base in the intergenic region, whereas the nucleotide following the locus number indicates the changed base. Information contained in the following three columns is similar to those described in Fig 2B. (C) Promoter region of *phuR* (region from -35 to +1) based on the PRODORIC database [84] is visualized with WebLogo [85], and SNPs listed in (B) are marked by asterisks. The height of a stack reflects the degree of sequence conservation at that position, while the height of symbols within a stack indicates the relative frequency of each nucleotide at that position. The upper sequence is derived from the intergenic regions of CF genomes and the sequence below is based on intergenic regions of non-CF genomes. (D) Red and black lines in the ‘id’ row indicate genomes of CF and non-CF isolates, respectively. Red lines in the bottom five rows indicate mutations at each locus, shown in (B), and yellow lines indicate that this intergenic region is missing from the genome. (E) Culture of *phuR::lacZ* and *phuRC117T::lacZ* grown to OD<sub>600nm</sub> ~0.25 were used in a β-galactosidase assay, to compare *phuR* promoter activity. \*\*\**p*<0.001.

<https://doi.org/10.1371/journal.ppat.1009681.g007>

results suggest that PA may benefit from small-scale mutations when establishing chronic infections in the CF airway. Mutations in the *lasR* gene encoding a QS regulator have been reported as frequent in CF isolates [7,8]. However, we found no evidence that mutations in LasR or other QS-related proteins were significantly overrepresented in CF or non-CF isolates. Given that the MucA mutation was represented in our analysis, demonstrating the precision of our bioinformatic approach, the *lasR* mutation might in fact be a common feature in PA, either causing chronic CF infection or other acute infections. Finally, while our study improves understanding the genetic changes in PA associated with chronic CF airway infection, they do not reveal the order in which these genetic changes occurred. Evaluating the chronology of adaptive changes in the CF lung will be important for explaining the persistence of PA.



**Fig 8. Schematic diagram with candidate genes.** Gene products of both (1–4) large- and (5–7) small-scale mutations significantly found in CF are illustrated with BioRender.com. We anticipate several phenotypic changes to arise as result of these mutations: (1) Large-scale mutations in the *MucA* protein stimulate alginate production which is an important component of biofilm; (2) PA2321 and PA3193 are involved in glucose catabolism. Disruption of PA2321 and PA3193 may be results of evolution in the CF environment which is enriched in free amino acids, nutrients preferred by *P. aeruginosa*; (3) Large-scale mutations in the PA2020 regulator upregulates expression level of MexXY, which may in turn result in increased antibiotic resistance; (4) Mutations in mannitol utilization gene cassette may aid in defence against the combined treatment of aminoglycoside and mannitol; (5) SLR deletion in PA5438 induces the glyoxylate shunt pathway, pyocyanin production and maintenance of biofilm in the presence of transferrin; (6) SLI insertion in *YecS* decreases cystine uptake and pyochelin production; (7) Mutation in *phuR* promoter region increases *phuR* expression.

<https://doi.org/10.1371/journal.ppat.1009681.g008>

## Experimental procedures

**Bioinformatic analysis.** Overall, 2,187 PA genomes were downloaded from the *Pseudomonas* Genome Database [27], and 1,001 genomes with information on the host disease were selected. A phylogenetic tree was created with RapidNJ, using Snippy-generated alignments of the core genome as input [67,68], and visualized with Microreact [69]. Based on the initial tree, the selection of genomes was trimmed down with Treemmer [70] to include 636 isolates while maintaining 99.8% of the original diversity (S1 Data). The prediction of protein-coding genes in the 636 genomes was performed using Prokka [31]. K-mers of 31 base pairs length (31-mers) were counted in the 636 individual genomes with Fsm-lite [71]. A similarity matrix was constructed with Snp-sim [72] using a core alignment of 636 genomes generated by Snippy [68]. Pyseer [29], which uses a mixed model (FaST-LMM), was run using the counted k-mers and the similarity matrix as input, and lrt p-values were assigned to each 31-mer. To determine the 31-mers that were significantly different between CF and non-CF groups, 31-mers with poor chi values (indicative of invalid  $\chi^2$  test) or lrt p-values above  $2.8e-08$  (cutoff value in this analysis) were removed. As a result, 41,685 31-mers were detected to have significant associations with either disease status (CF or non-CF). Distribution of these 31-mers aligned to the PAO1 whole genome was visualized with Phandango [30]. Subsequently, *de*

*novo* assembly of these 31-mers was completed using Trinity [73]. Overall, 489 contigs were constructed, and these contigs mapped to 29 intergenic regions, six untranslated regions, and 109 translated genes of the PAO1 genome with Blastn. Individual contigs that aligned to either protein-coding regions or intergenic regions were blasted against 635 genomes (the PAO1 genome was excluded) [74]. The top hit from each pairwise alignment with a blast e-value <0.01 was chosen as a homolog of the candidate sequence in each isolate.

In 453 contigs mapping to protein-coding regions, homologs were clustered under several conditions of similarity and gene length coverage (S95L95 -: 95% similarity and 95% length coverage; S70L95-70% similarity and 95% length coverage) using Blastclust [74]. Homologs that clustered with the reference PAO1 protein were assigned as '1' and those that did not were assigned '0'. We then used the clustering matrix as input for Pyseer. Clusters with either poor chi values or lrt p-values larger than 3.22E-05 were removed. In this manner, genes with multiple co-localized mutations were determined and defined as 'large-scale' mutations (Fig 1C). Other protein-coding region mutations were treated as 'small-scale' mutations.

Where contigs mapped to protein-coding regions or intergenic regions, multiple alignment of the candidate sequence and its homologs was performed using Mafft [75]. This allowed characterization of nucleotides (for intergenic region candidates) or amino acid sequences (for protein-coding genes) at each locus for detecting small-scale mutations. If the residue of a homolog was identical to that of the reference PAO1 intergenic region or protein, it was assigned a value of '1' at that locus, and if not, a value of '0' was assigned. In this way, a locus matrix was created for use as the input for Pyseer and loci with either poor chi values or lrt p-values larger than 4.01E-06 were removed.

Finally, k-mers that failed to align to the PAO1 genome were also analyzed. However, the PAO1 reference gene in these cases was replaced with the reference gene detected in alternative isolate genomes. As such, we were able to investigate the presence of insertion mutations. For KEGG functional annotation and pathway analysis, BlastKoala [34] was used with large-scale mutations. Sequence comparison was performed by Clustalw [76] and multiple alignment was visualized with Jalview [77]. SUPERFAMILY 2 [41] search was used to investigate the domains of the PA5438 gene. Transmembrane domain prediction of the AU17965\_3981\_04951 protein was carried out using Phobius [56].

**Bacterial strains and growth conditions.** All bacterial strains and plasmids used in this study are shown in Table C in S1 Table. PAO1 was used as a reference strain [1]. Bacterial cultures were grown in LB medium (1% [w/v] tryptone, 0.5% [w/v] yeast extract, and 1% [w/v] sodium chloride) at 37°C and Synthetic Cystic Fibrosis Medium (SCFM) using a previously described protocol [59]. Single bacterial colonies were picked from LB plates and inoculated in fresh LB or SCFM broths for precultures and grown overnight. Precultures were diluted 100-fold in fresh LB broth to subculture and incubated at 37°C with shaking at 230 rpm. For SCFM cultures, bacterial preculture was diluted and inoculated to the starting OD<sub>600nm</sub> 0.025 in fresh SCFM broth for subculture. All SCFM cultures were incubated under the same conditions as LB cultures, except for the incubation time, which was dependent on specific experimental procedures.

### In-frame mutant construction

All in-frame mutants, including an SLI insertion in the YecS protein and SLR deletion in the PA5438 protein, were constructed from the PAO1 strain. In-frame insertions and deletions were performed to construct mutants with amino acid level changes. In the case of the in-frame deletions, the 5' and 3' flanking regions were designed to overlap in both directions of SLR. For the in-frame insertion, nucleotide sequences corresponding to amino acid sequence 'SLI' were

inserted into the middle of the 5' flanking region. An overlap was constructed by using each 5' primer of 5' flanking region and 3' primer of 3' flanking region. For both deletion and insertion mutations, an overlapping product was inserted into modified pCVD442, a suicide vector, containing gentamicin and ampicillin resistance markers. PAO1, grown on LB agar, was conjugated with *E. coli* SM10  $\lambda$ pir, harboring pCVD442 with the overlapping product inserted, grown on LB agar with 50  $\mu$ g/ml ampicillin and 30  $\mu$ g/ml gentamicin. Conjugates were spread onto LB agar with 50  $\mu$ g/ml gentamicin and 20  $\mu$ g/ml irgasan to select single crossover recombinants. The single crossover recombinant was incubated on LB agar without NaCl but containing 8% sucrose for the selection of the desired mutant. Sequence verification was performed by PCR. Primers used in constructing these mutants are listed in Table D in [S1 Table](#).

## Complementation

Identical primers were used in constructing the aforementioned mutants (Table D in [S1 Table](#); PA5438#1, PA5438#4, *yecS*#1, and *yecS*#6), and were used to amplify the *yecS* and PA5438 sequences of wildtype PAO1. The amplified wildtype sequences were each inserted into pCVD442, and identical steps of mutant construction were followed to generate the complementation strains of the Yec $\Omega$ SLI and PA5438 $\Delta$ SLR mutants.

## Promoter assay

To perform the *phuR* gene promoter assay, intergenic regions (180 bp) between *phuS* and *phuR* in the PAO1 genome and AU2342\_3932 [78] with only the 117<sup>th</sup> locus changed from cytosine to thymine were amplified with primers listed in Table D in [S1 Table](#). These PCR-products were each cloned into the upstream region of the  $\beta$ -galactosidase gene of puc18-mini-Tn7t-Gm-LacZ [79] for chromosomal insertion. The constructed plasmid was transformed into *E. coli* DH5 $\alpha$   $\lambda$ pir. After verification by DNA sequencing, the plasmids with the helper plasmid pTNS2 that encodes TnsABC+D genes (allowing Tn7 transposition) were electroporated into PAO1. Empty puc18-mini-Tn7t-Gm-LacZ plasmid with no insert was used as a control to measure the baseline expression of *lacZ*. The potential clones were selected on LB agar with 50  $\mu$ g/ml gentamicin and sequence verification was performed to select the final candidates whose transposon was inserted properly into the region following the *glmS* gene.  $\beta$ -galactosidase activities of the three clones (*con::lacZ*, *phuR::lacZ*, and *phuRC117T::lacZ*), summarized in Table C in [S1 Table](#), were grown in LB and measured at the exponential phase (OD<sub>600nm</sub> ~0.25).

## Growth curves

Wildtype PAO1, the PA5438 $\Delta$ SLR mutant, and the complementation strain were precultured overnight in LB broth. Precultures were diluted 100-fold in fresh LB and incubated at 37°C with shaking at 230 rpm. Growth in LB was observed over 12 hours and the OD<sub>600nm</sub> was measured. Growth in SCFM was detected in the same manner over a period of 8 hours. Growth recovery tests of the PA5438 $\Delta$ SLR mutant were performed using LB medium and LB supplemented with 10 mM sodium acetate (Sigma-Aldrich, USA), 10 mM sodium pyruvate (Sigma-Aldrich, USA), and 10 mM sodium citrate (Sigma-Aldrich, USA), and OD<sub>600nm</sub> were measured at regular time intervals. To examine the functional capacity of the L-cystine transporter, overnight precultures of PAO1, the Yec $\Omega$ SLI mutant, and the complementation strain were washed in phosphate-buffered saline (PBS), and diluted to OD<sub>600nm</sub> ~ 0.025 in fresh M9 minimal media supplemented with 22.2 mM glucose and L-cystine (0.25 mM) as the sole sulfur source (2 mM MgCl<sub>2</sub> instead of MgSO<sub>4</sub>). OD<sub>600nm</sub> was measured over 30 hours. Growth was

also measured over 21 hours under identical conditions, except for the inclusion of 2 mM MgSO<sub>4</sub> (instead of MgCl<sub>2</sub>) in M9 minimal media.

### Antibiotic susceptibility test

For antibiotic susceptibility tests, ciprofloxacin (Duchefa, The Netherlands) and tobramycin (Sigma-Aldrich, USA) were used. Wildtype PAO1, the PA5438ΔSLR mutant, and the complementation strain were precultured overnight in LB broth. Bacterial preculture was diluted in LB containing the relevant antibiotics (ciprofloxacin or tobramycin) and the starting OD<sub>600nm</sub> was adjusted to 0.005. For static cultures, the antibiotic concentrations used were 0.125 μg/ml ciprofloxacin and 1 μg/ml tobramycin. OD<sub>600nm</sub> was measured after 20 and 26 hours of static incubation with ciprofloxacin and tobramycin, respectively. An antibiotic susceptibility test with 10 mM sodium acetate supplementation was conducted under the same conditions.

### Reverse transcription and quantitative real-time PCR

PAO1, the PA5438ΔSLR mutant, and the complementation strain were precultured and subcultured in LB. After incubating the subcultures to OD<sub>600nm</sub> ~1.0, RNeasy Mini kit (Qiagen, Netherland) and on-column DNaseI digestion were used to extract RNA following the manufacturer's protocol. 1 μg of RNA was reverse-transcribed to synthesize complementary DNA by using reverse transcriptase (Takara Bio, Japan) and random hexamer primers. To check for DNA contamination, the same cDNA synthesis was performed in the absence of reverse transcriptase. SYBR green-based qPCR was performed using an ABI 48-well StepOne real-time system. The primers used are listed in Table D in [S1 Table](#). CT values were normalized by 16S rRNA CT values. SCFM cultures (OD<sub>600nm</sub> ~0.9) followed identical RNA extraction and qRT-PCR procedures.

### Elastase and pyocyanin tests

Precultures of PAO1, the PA5438ΔSLR mutant, and the complementation strain were prepared for elastase and pyocyanin tests in SCFM as described above. Bacterial supernatant was harvested from bacterial subcultures incubated for 8 hours in SCFM, and assays were performed as described previously [80,81]. Elastase and pyocyanin values were normalized using OD<sub>600nm</sub>.

### Pyochelin measurement

Precultures of PAO1, the YecSΩSLI mutant, and the complementation strain in SCFM were prepared as described above. Bacterial supernatant was harvested from bacterial subcultures, incubated for 6 hours in SCFM, and fluorescence was measured using Varioskan Flash 3001 (Thermo Scientific, USA) at excitation wavelength 355 nm and emission wavelength 440 nm [82]. Pyochelin values were normalized using OD<sub>600nm</sub>.

### Biofilm assay

Precultures of PAO1, PA5438ΔSLR mutant, and the complementation strain in SCFM were prepared as described above. Bacterial precultures were diluted in SCFM or SCFM containing the holo-transferrin (Sigma-Aldrich, USA) in a 96-well plate, and the starting OD<sub>600nm</sub> was adjusted to 0.025. For static cultures, the holo-transferrin concentration used was 1.25 μM. After 32 hours of static incubation, biofilm assay was performed as described previously [83].

### In vitro gentamicin protection assay

Neutrophils were isolated from the bone-marrow of C57BL/6 mice following the protocol of EasySep Mouse Neutrophil Enrichment Kit (STEMCELL, Canada). Following isolation, neutrophils were maintained in Opti-MEM media (ThermoFisher Scientific, USA). Precultures of PAO1, the PA5438ΔSLR mutant, and the complementation strain in SCFM were prepared in the manner described above. Bacterial culture was harvested from bacterial subcultures incubated for 4 hours in SCFM. The initial infection dosage was  $10^7$  CFU per  $5 \times 10^5$  neutrophils. After 2 hours of co-culturing *P. aeruginosa* and neutrophils, the culture supernatant was removed and fresh Opti-MEM supplemented with 50 μg/ml gentamicin was added for 1 hour. After performing wash twice in PBS, 0.5% Triton-X was added and bacteria number were counted by using pseudomonas isolation agar (Sigma-Aldrich, USA).

### Statistical analysis

Data are expressed as mean ± standard deviation. Unpaired Student's t-test (one-tailed, unequal variance) was performed to analyze the differences between experimental groups. P-values smaller than 0.05 were considered statistically significant. All experiments were repeated for reproducibility.

### Supporting information

**S1 Fig. Phylogenetic tree constructed with CF and non-CF genomes.** Upper phylogenetic tree was constructed with 1,001 genomes containing host disease information, and the tree below was drawn using 636 genomes and maintaining 99.8% diversity of the upper tree. Black and red leaves each indicate non-CF and CF isolates, and the numbers of CF and non-CF genomes for constructing each phylogenetic tree are placed inside brackets. (TIF)

**S2 Fig. Phenotypes of the PA5438ΔSLR mutant.** (A) Antibiotic susceptibility test with tobramycin (Tob) and Tob with 10 mM sodium acetate (SA) were performed. Initial  $OD_{600nm}$  of PAO1, PA5438ΔSLR mutant, and the complementation strain were adjusted to 0.005, and  $OD_{600nm}$  was measured after 26 hours of static incubation in LB supplemented with Tob. The concentration of antibiotic was 1 μg/ml. \*\* $p < 0.01$  (B) Growth curves of PAO1 and PA5438ΔSLR mutant in SCFM were observed over 8 hours. (C) RNAs of PAO1, the PA5438ΔSLR mutant, and the complementation strain were extracted at  $OD_{600nm} \sim 0.9$  in SCFM and relative expression levels of *aceE* were measured. (D) Relative expression levels of *aceA* and *glcB* of the same RNA used in (C) were measured. \*\*\* $p < 0.001$  (E) The initial infection dosage was  $10^7$  CFU per  $5 \times 10^5$  neutrophils isolated from the bone-marrow of C57BL/6 mice. After 2 hours of co-culturing *P. aeruginosa* and neutrophils, we added gentamicin to the culture medium for 1 hour to remove any extracellular bacteria. We then harvested the intracellular bacteria by treating the neutrophils with 0.5% Triton-X, and measured the bacterial CFU. (F) After bacterial culture of PAO1, PA5438ΔSLR mutant, and the complementation strain in SCFM for 8 hours with shaking, an elastase assay was performed with culture supernatant. \*\* $p < 0.01$ . (TIF)

**S3 Fig. Blastp results of genes associated with pyocyanin biosynthesis.** Each row represents a genome containing an SLR deletion in PA5438 homologs and color indicates whether it is included in CF or non-CF. Columns contain PAO1 proteins associated with pyocyanin biosynthesis and similarity (S), % of identical matches between reference PAO1 protein and its homolog, and length coverage (L), % of reference PAO1 protein sequence covered by its

homolog. Gray box indicates there is no homolog under e-value 0.01 in the blastp search. (TIF)

**S4 Fig. Multiple alignment of YecS and its homologs.** Multiple alignment of YecS and its homologs is visually represented. Names of the genomes are shown to the left of the multiple alignment. PAO1 and AU17965\_3981 are representative genomes of non-CF and CF groups, and genomes labeled with red or blue belong to the CF or non-CF groups, respectively. Black arrow marks where the additional SLI insertion occurs (162<sup>nd</sup> to 164<sup>th</sup> residues) compared to the YecS protein. Regions marked by asterisks are regions of SLI amino acid repeat sequences in YecS. Genomes in **a** contain a deletion of SLI, resulting in a single copy of SLI, and genomes in **b** present large deletions in YecS. (TIF)

**S5 Fig. Transmembrane domain prediction and growth curves of the YecS mutant.** (A) Predicted transmembrane domains of the AU17965\_3981\_04951 protein are portrayed. Numbers below the figure indicate the amino acid loci. (B) Detailed amino acid ranges of the predicted transmembrane domains are listed. The SLI insertion in the AU17965\_3981\_4951 protein is present within the region highlighted in bold. (C) Growth curves of PAO1 and YecS $\Omega$ SLI mutant in SCFM were measured over 8 hours. (D) Growth curves of PAO1 and YecS $\Omega$ SLI in M9 minimal media supplemented with glucose, L-cystine, and MgSO<sub>4</sub> were recorded over 21 hours. (TIF)

**S6 Fig. Blastp results of genes associated with pyochelin biosynthesis.** Each row represents a genome containing an SLI insertion in YecS homologs and color indicates whether it is included in CF or non-CF. Columns contain PAO1 proteins associated with pyochelin biosynthesis and similarity (S), % of identical matches between reference PAO1 protein and its homolog, and length coverage (L), % of reference PAO1 protein sequence covered by its homolog. Gray box indicates there is no homolog under e-value 0.01 in the blastp search. (TIF)

**S1 Data. Information of 636 genomes.**

(XLSX)

**S2 Data. Pyseer results.**

(XLSX)

**S3 Data. Cluster and locus information of 636 genomes.**

(XLSX)

**S4 Data. Amino acid sequences of small insertion variants from isolates other than PAO1.**

(DOCX)

**S5 Data. Blastp results with 636 genomes and 238 YdjN homologs.**

(XLSX)

**S1 Table. Supplementary tables.**

(DOCX)

## Author Contributions

**Conceptualization:** Wontae Hwang, Sang Sun Yoon.

**Data curation:** Wontae Hwang, Ji Hyun Yong, Kyung Bae Min, Kang-Mu Lee.

**Formal analysis:** Wontae Hwang.

**Funding acquisition:** Sang Sun Yoon.

**Investigation:** Wontae Hwang.

**Methodology:** Wontae Hwang, Kyung Bae Min, Kang-Mu Lee.

**Software:** Ben Pascoe, Samuel K Sheppard.

**Supervision:** Sang Sun Yoon.

**Validation:** Wontae Hwang, Ji Hyun Yong.

**Visualization:** Wontae Hwang.

**Writing – original draft:** Wontae Hwang, Ji Hyun Yong.

**Writing – review & editing:** Wontae Hwang, Ji Hyun Yong, Ben Pascoe, Samuel K Sheppard, Sang Sun Yoon.

## References

1. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warriner P, Hickey MJ, et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*. 2000; 406(6799):959–64. <https://doi.org/10.1038/35023079> PMID: 10984043
2. Bhagirath AY, Li Y, Somayajula D, Dadashi M, Badr S, Duan K. Cystic fibrosis lung environment and *Pseudomonas aeruginosa* infection. *BMC Pulm Med*. 2016; 16(1):174-. <https://doi.org/10.1186/s12890-016-0339-5> PMID: 27919253.
3. Silby MW, Winstanley C, Godfrey SAC, Levy SB, Jackson RW. *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiology Reviews*. 2011; 35(4):652–80. <https://doi.org/10.1111/j.1574-6976.2011.00269.x> PMID: 21361996
4. Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, D'Argenio DA, et al. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A*. 2006; 103(22):8487–92. Epub 2006/05/10. <https://doi.org/10.1073/pnas.0602138103> PMID: 16687478.
5. Palmer KL, Mashburn LM, Singh PK, Whiteley M. Cystic fibrosis sputum supports growth and cues key aspects of *Pseudomonas aeruginosa* physiology. *J Bacteriol*. 2005; 187(15):5267–77. <https://doi.org/10.1128/JB.187.15.5267-5277.2005> PMID: 16030221.
6. Fung C, Naughton S, Turnbull L, Tingpej P, Rose B, Arthur J, et al. Gene expression of *Pseudomonas aeruginosa* in a mucin-containing synthetic growth medium mimicking cystic fibrosis lung sputum. *J Med Microbiol*. 2010; 59(Pt 9):1089–100. <https://doi.org/10.1099/jmm.0.019984-0> PMID: 20522626.
7. Hoffman LR, Kulasekara HD, Emerson J, Houston LS, Burns JL, Ramsey BW, et al. *Pseudomonas aeruginosa* lasR mutants are associated with cystic fibrosis lung disease progression. *J Cyst Fibros*. 2009; 8(1):66–70. Epub 2008/11/01. <https://doi.org/10.1016/j.jcf.2008.09.006> PMID: 18974024.
8. Feltner JB, Wolter DJ, Pope CE, Groleau MC, Smalley NE, Greenberg EP, et al. LasR Variant Cystic Fibrosis Isolates Reveal an Adaptable Quorum-Sensing Hierarchy in *Pseudomonas aeruginosa*. *mBio*. 2016; 7(5). Epub 2016/10/06. <https://doi.org/10.1128/mBio.01513-16> PMID: 27703072.
9. Salunkhe P, Smart CH, Morgan JA, Panagea S, Walshaw MJ, Hart CA, et al. A cystic fibrosis epidemic strain of *Pseudomonas aeruginosa* displays enhanced virulence and antimicrobial resistance. *J Bacteriol*. 2005; 187(14):4908–20. Epub 2005/07/05. <https://doi.org/10.1128/JB.187.14.4908-4920.2005> PMID: 15995206.
10. Mustafa MH, Khandekar S, Tunney MM, Elborn JS, Kahl BC, Denis O, et al. Acquired resistance to macrolides in *Pseudomonas aeruginosa* from cystic fibrosis patients. *Eur Respir J*. 2017; 49(5). Epub 2017/05/21. <https://doi.org/10.1183/13993003.01847-2016> PMID: 28526799.
11. Maunders EA, Triniman RC, Western J, Rahman T, Welch M. Global reprogramming of virulence and antibiotic resistance in *Pseudomonas aeruginosa* by a single nucleotide polymorphism in elongation factor, fusA1. *J Biol Chem*. 2020. Epub 2020/09/19. <https://doi.org/10.1074/jbc.RA119.012102> PMID: 32943550.
12. Barth AL, Pitt TL. Auxotrophic variants of *Pseudomonas aeruginosa* are selected from prototrophic wild-type strains in respiratory infections in patients with cystic fibrosis. *J Clin Microbiol*. 1995; 33(1):37–40. <https://doi.org/10.1128/jcm.33.1.37-40.1995> PMID: 7699062.

13. Barth A, Pitt TL. The high amino-acid content of sputum from cystic fibrosis patients promotes growth of auxotrophic *Pseudomonas aeruginosa*. *J Med Microbiol*. 1996; 45:110–9. <https://doi.org/10.1099/00222615-45-2-110> PMID: 8683546
14. Marvig RL, Sommer LM, Molin S, Johansen HK. Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. *Nature Genetics*. 2015; 47(1):57–64. <https://doi.org/10.1038/ng.3148> PMID: 25401299
15. Quinn RA, Lim YW, Mak TD, Whiteson K, Furlan M, Conrad D, et al. Metabolomics of pulmonary exacerbations reveals the personalized nature of cystic fibrosis disease. *PeerJ*. 2016; 4:e2174–e. <https://doi.org/10.7717/peerj.2174> PMID: 27602256.
16. Diaz Caballero J, Clark ST, Coburn B, Zhang Y, Wang PW, Donaldson SL, et al. Selective Sweeps and Parallel Pathoadaptation Drive *Pseudomonas aeruginosa* Evolution in the Cystic Fibrosis Lung. *mBio*. 2015; 6(5):e00981. <https://doi.org/10.1128/mBio.00981-15> PMID: 26330513.
17. La Rosa R, Johansen HK, Molin S. Convergent Metabolic Specialization through Distinct Evolutionary Paths in *Pseudomonas aeruginosa*. *mBio*. 2018; 9(2):e00269–18. <https://doi.org/10.1128/mBio.00269-18> PMID: 29636437.
18. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. *Nat Rev Genet*. 2010; 11(5):356–66. Epub 2010/04/17. <https://doi.org/10.1038/nrg2760> PMID: 20395969.
19. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019; 47(D1):D1005–D12. <https://doi.org/10.1093/nar/gky1120> PMID: 30445434.
20. Hwang W, Yoon SS. Virulence Characteristics and an Action Mode of Antibiotic Resistance in Multi-drug-Resistant *Pseudomonas aeruginosa*. *Sci Rep*. 2019; 9(1):487. Epub 2019/01/27. <https://doi.org/10.1038/s41598-018-37422-9> PMID: 30679735.
21. Chen PE, Shapiro BJ. The advent of genome-wide association studies for bacteria. *Current Opinion in Microbiology*. 2015; 25:17–24. <https://doi.org/10.1016/j.mib.2015.03.002> PMID: 25835153
22. Lees JA, Vehkala M, Välimäki N, Harris SR, Chewapreecha C, Croucher NJ, et al. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nature Communications*. 2016; 7(1):12797. <https://doi.org/10.1038/ncomms12797> PMID: 27633831
23. Sheppard SK, Didelot X, Meric G, Torralbo A, Jolley KA, Kelly DJ, et al. Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in *Campylobacter*. *Proc Natl Acad Sci U S A*. 2013; 110(29):11923–7. Epub 2013/07/01. <https://doi.org/10.1073/pnas.1305559110> PMID: 23818615.
24. Sheppard SK, Guttman DS, Fitzgerald JR. Population genomics of bacterial host adaptation. *Nature Reviews Genetics*. 2018; 19(9):549–65. <https://doi.org/10.1038/s41576-018-0032-z> PMID: 29973680
25. Earle SG, Wu C-H, Charlesworth J, Stoesser N, Gordon NC, Walker TM, et al. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nat Microbiol*. 2016; 1:16041-. <https://doi.org/10.1038/nmicrobiol.2016.41> PMID: 27572646.
26. Jaillard M, Lima L, Tournoud M, Mahe 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; 14(11):e1007758. Epub 2018/11/13. <https://doi.org/10.1371/journal.pgen.1007758> PMID: 30419019 following competing interests: MJ, MT, PM and AvB are employees of bioMerieux, a company that develops and sells diagnostic tests in the field of infectious diseases. However, the study was designed and executed in an open manner and the presented method as well as all data generated have been deposited in the public domain, also resulting in the current publication.
27. Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FSL. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res*. 2016; 44(D1):D646–D53. Epub 2015/11/17. <https://doi.org/10.1093/nar/gkv1227> PMID: 26578582.
28. Ozer EA, Nnah E, Didelot X, Whitaker RJ, Hauser AR. The Population Structure of *Pseudomonas aeruginosa* Is Characterized by Genetic Isolation of exoU+ and exoS+ Lineages. *Genome Biol Evol*. 2019; 11(1):1780–96. <https://doi.org/10.1093/gbe/evz119> PMID: 31173069.
29. Lees JA, Galardini M, Bentley SD, Weiser JN, Corander J. pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics*. 2018; 34(24):4310–2. <https://doi.org/10.1093/bioinformatics/bty539> PMID: 30535304
30. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics*. 2018; 34(2):292–3. <https://doi.org/10.1093/bioinformatics/btx610> PMID: 29028899

31. Seemann T, Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014; 30(14):2068–9. <https://doi.org/10.1093/bioinformatics/btu153> PMID: 24642063
32. Anthony M, Rose B, Pegler MB, Elkins M, Service H, Thamotharampillai K, et al. Genetic analysis of *Pseudomonas aeruginosa* isolates from the sputa of Australian adult cystic fibrosis patients. *J Clin Microbiol*. 2002; 40(8):2772–8. <https://doi.org/10.1128/JCM.40.8.2772-2778.2002> PMID: 12149328.
33. Boucher JC, Yu H, Mudd M, Deretic V. Mucoid *Pseudomonas aeruginosa* in cystic fibrosis: Characterization of muc mutations in clinical isolates and analysis of clearance in a mouse model of respiratory infection. *Infection and immunity*. 1997; 65:3838–46. <https://doi.org/10.1128/iai.65.9.3838-3846.1997> PMID: 9284161
34. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*. 2016; 428(4):726–31. <https://doi.org/10.1016/j.jmb.2015.11.006> PMID: 26585406
35. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000; 28(1):27–30. <https://doi.org/10.1093/nar/28.1.27> PMID: 10592173.
36. Yoon SS, Coakley R, Lau GW, Lyman SV, Gaston B, Karabulut AC, et al. Anaerobic killing of mucoid *Pseudomonas aeruginosa* by acidified nitrite derivatives under cystic fibrosis airway conditions. *J Clin Invest*. 2006; 116(2):436–46. Epub 2006/01/26. <https://doi.org/10.1172/JCI24684> PMID: 16440061.
37. Min KB, Lee K-M, Oh YT, Yoon SS. Nonmucoid conversion of mucoid *Pseudomonas aeruginosa* induced by sulfate-stimulated growth. *FEMS Microbiology Letters*. 2014; 360(2):157–66. <https://doi.org/10.1111/1574-6968.12600> PMID: 25227776
38. Morita Y, Tomida J, Kawamura Y. MexXY multidrug efflux system of *Pseudomonas aeruginosa*. *Front Microbiol*. 2012; 3:408-. <https://doi.org/10.3389/fmicb.2012.00408> PMID: 23233851.
39. Rojo F. Carbon catabolite repression in *Pseudomonas*: optimizing metabolic versatility and interactions with the environment. *FEMS Microbiology Reviews*. 2010; 34(5):658–84. <https://doi.org/10.1111/j.1574-6976.2010.00218.x> PMID: 20412307
40. Barraud N, Buson A, Jarolimek W, Rice SA. Mannitol enhances antibiotic sensitivity of persister bacteria in *Pseudomonas aeruginosa* biofilms. *PLoS One*. 2013; 8(12):e84220–e. <https://doi.org/10.1371/journal.pone.0084220> PMID: 24349568.
41. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. Edited by G. Von Heijne. *Journal of Molecular Biology*. 2001; 313(4):903–19. <https://doi.org/10.1006/jmbi.2001.5080> PMID: 11697912
42. Dolan SK, Pereira G, Silva-Rocha R, Welch M. Transcriptional regulation of central carbon metabolism in *Pseudomonas aeruginosa*. *Microb Biotechnol*. 2020; 13(1):285–9. Epub 2019/06/11. <https://doi.org/10.1111/1751-7915.13423> PMID: 31187593.
43. Campilongo R, Fung RKY, Little RH, Grenga L, Trampari E, Pepe S, et al. One ligand, two regulators and three binding sites: How KDPG controls primary carbon metabolism in *Pseudomonas*. *PLOS Genetics*. 2017; 13(6):e1006839. <https://doi.org/10.1371/journal.pgen.1006839> PMID: 28658302
44. Ciofu O, Tolker-Nielsen T. Tolerance and Resistance of *Pseudomonas aeruginosa* Biofilms to Antimicrobial Agents—How *P. aeruginosa* Can Escape Antibiotics. *Front Microbiol*. 2019; 10:913-. <https://doi.org/10.3389/fmicb.2019.00913> PMID: 31130925.
45. Liu GY, Nizet V. Color me bad: microbial pigments as virulence factors. *Trends Microbiol*. 2009; 17(9):406–13. Epub 2009/08/31. <https://doi.org/10.1016/j.tim.2009.06.006> PMID: 19726196.
46. Cox CD. Role of pyocyanin in the acquisition of iron from transferrin. *Infection and immunity*. 1986; 52(1):263–70. <https://doi.org/10.1128/iai.52.1.263-270.1986> PMID: 2937736.
47. Ghio AJ, Roggli VL, Soukup JM, Richards JH, Randell SH, Muhlebach MS. Iron accumulates in the lavage and explanted lungs of cystic fibrosis patients. *Journal of Cystic Fibrosis*. 2013; 12(4):390–8. <https://doi.org/10.1016/j.jcf.2012.10.010> PMID: 23176785
48. Meyer KC, Sharma A, Brown R, Weatherly M, Moya FR, Lewandoski J, et al. Function and Composition of Pulmonary Surfactant and Surfactant-Derived Fatty Acid Profiles Are Altered in Young Adults With Cystic Fibrosis. *Chest*. 2000; 118(1):164–74. <https://doi.org/10.1378/chest.118.1.164> PMID: 10893374
49. Renilla S, Bernal V, Fuhrer T, Castaño-Cerezo S, Pastor JM, Iborra JL, et al. Acetate scavenging activity in *Escherichia coli*: interplay of acetyl-CoA synthetase and the PEP-glyoxylate cycle in chemostat cultures. *Applied Microbiology and Biotechnology*. 2012; 93(5):2109–24. <https://doi.org/10.1007/s00253-011-3536-4> PMID: 21881893
50. Son MS, Matthews WJ Jr, Kang Y, Nguyen DT, Hoang TT. In vivo evidence of *Pseudomonas aeruginosa* nutrient acquisition and pathogenesis in the lungs of cystic fibrosis patients. *Infection and immunity*. 2007; 75(11):5313–24. Epub 2007/08/27. <https://doi.org/10.1128/IAI.01807-06> PMID: 17724070.

51. Flynn JM, Phan C, Hunter RC. Genome-Wide Survey of *Pseudomonas aeruginosa* PA14 Reveals a Role for the Glyoxylate Pathway and Extracellular Proteases in the Utilization of Mucin. *Infection and immunity*. 2017; 85(8):e00182–17. <https://doi.org/10.1128/IAI.00182-17> PMID: 28507068.
52. Rossi E, Falcone M, Molin S, Johansen HK. High-resolution in situ transcriptomics of *Pseudomonas aeruginosa* unveils genotype independent patho-phenotypes in cystic fibrosis lungs. *Nature Communications*. 2018; 9(1):3459. <https://doi.org/10.1038/s41467-018-05944-5> PMID: 30150613
53. Lorenz MC, Fink GR. Life and death in a macrophage: role of the glyoxylate cycle in virulence. *Eukaryot Cell*. 2002; 1(5):657–62. <https://doi.org/10.1128/EC.1.5.657-662.2002> PMID: 12455685.
54. Lemire J, Alhasawi A, Appanna VP, Tharmalingam S, Appanna VD. Metabolic defence against oxidative stress: the road less travelled so far. *Journal of Applied Microbiology*. 2017; 123(4):798–809. <https://doi.org/10.1111/jam.13509> PMID: 28609580
55. Donlan RM. Biofilms: microbial life on surfaces. *Emerg Infect Dis*. 2002; 8(9):881–90. <https://doi.org/10.3201/eid0809.020063> PMID: 12194761.
56. Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res*. 2007; 35(Web Server issue):W429–W32. Epub 2007/05/05. PMID: 17483518.
57. Ohtsu I, Kawano Y, Suzuki M, Morigasaki S, Saiki K, Yamazaki S, et al. Uptake of L-cystine via an ABC transporter contributes defense of oxidative stress in the L-cystine export-dependent manner in *Escherichia coli*. *PLoS One*. 2015; 10(3):e0120619–e. <https://doi.org/10.1371/journal.pone.0120619> PMID: 25837721.
58. The UniProt C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019; 47(D1):D506–D15. PMID: 30395287
59. Palmer KL, Aye LM, Whiteley M. Nutritional cues control *Pseudomonas aeruginosa* multicellular behavior in cystic fibrosis sputum. *J Bacteriol*. 2007; 189(22):8079–87. Epub 2007/09/14. <https://doi.org/10.1128/JB.01138-07> PMID: 17873029.
60. Korshunov S, Imlay KRC, Imlay JA. Cystine import is a valuable but risky process whose hazards *Escherichia coli* minimizes by inducing a cysteine exporter. *Molecular Microbiology*. 2020; 113(1):22–39. <https://doi.org/10.1111/mmi.14403> PMID: 31612555
61. Cornelis P, Dingemans J. *Pseudomonas aeruginosa* adapts its iron uptake strategies in function of the type of infections. *Front Cell Infect Microbiol*. 2013; 3:75-. <https://doi.org/10.3389/fcimb.2013.00075> PMID: 24294593.
62. Ronnebaum TA, Lamb AL. Nonribosomal peptides for iron acquisition: pyochelin biosynthesis as a case study. *Current Opinion in Structural Biology*. 2018; 53:1–11. <https://doi.org/10.1016/j.sbi.2018.01.015> PMID: 29455106
63. Lee K, Lee K-M, Go J, Ryu J-C, Ryu J-H, Yoon SS. The ferrichrome receptor A as a new target for *Pseudomonas aeruginosa* virulence attenuation. *FEMS Microbiology Letters*. 2016; 363(11). PMID: 27190289
64. Ochsner U, Johnson Z, Vasil M. Genetics and regulation of two distinct haem-uptake systems, phu and has, in *Pseudomonas aeruginosa*. *Microbiology (Reading, England)*. 2000; 146 (Pt 1):185–98. <https://doi.org/10.1099/00221287-146-1-185> PMID: 10658665
65. Marvig RL, Damkjaer S, Khademi SMH, Markussen TM, Molin S, Jelsbak L. Within-host evolution of *Pseudomonas aeruginosa* reveals adaptation toward iron acquisition from hemoglobin. *mBio*. 2014; 5 (3):e00966. <https://doi.org/10.1128/mBio.00966-14> PMID: 24803516.
66. Kobras CM, Fenton AK, Sheppard SK. Next-generation microbiology: from comparative genomics to gene function. *Genome Biol*. 2021; 22(1):123-. <https://doi.org/10.1186/s13059-021-02344-9> PMID: 33926534.
67. Simonsen M, Mailund T, Pedersen CNS, editors. Rapid Neighbour-Joining. *Algorithms in Bioinformatics*; 2008 2008//; Berlin, Heidelberg: Springer Berlin Heidelberg.
68. Seemann T. Snippy. <https://github.com/tseemann/snippy>.
69. 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; 2(11):e000093–e. <https://doi.org/10.1099/mgen.0.000093> PMID: 28348833.
70. Menardo F, Loiseau C, Brites D, Coscolla M, Gygli SM, Rutaiwa LK, et al. Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity. *BMC Bioinformatics*. 2018; 19(1):164. <https://doi.org/10.1186/s12859-018-2164-8> PMID: 29716518
71. Välimäki N. Fsm-lite. <https://github.com/nvalimak/fsm-lite>.
72. Bayliss S. Snp-sim. <https://github.com/SionBayliss/bio-scripts/tree/master/scripts>.

73. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011; 29(7):644–52. <https://doi.org/10.1038/nbt.1883> PMID: 21572440.
74. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC bioinformatics*. 2009; 10:421–. <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500.
75. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30(4):772–80. Epub 2013/01/16. PMID: 23329690.
76. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994; 22(22):4673–80. PMID: 7984417.
77. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)*. 2009; 25(9):1189–91. Epub 2009/01/16. PMID: 19151095.
78. Spilker T, LiPuma JJ. Draft Genome Sequences of 63 *Pseudomonas aeruginosa* Isolates Recovered from Cystic Fibrosis Sputum. *Genome Announc*. 2016; 4(2):e00231–16. <https://doi.org/10.1128/genomeA.00231-16> PMID: 27103710.
79. Choi K-H, Gaynor JB, White KG, Lopez C, Bosio CM, Karkhoff-Schweizer RR, et al. A Tn7-based broad-range bacterial cloning and expression system. *Nature Methods*. 2005; 2(6):443–8. <https://doi.org/10.1038/nmeth765> PMID: 15908923
80. Sezgi S, Seyhan U, Gulgun B-T, Aysegul Y. Quorum sensing and virulence of *Pseudomonas aeruginosa* during urinary tract infections. *The Journal of Infection in Developing Countries*. 2012; 6(06). <https://doi.org/10.3855/jidc.2543> PMID: 22706193
81. Carlsson M, Shukla S, Petersson AC, Segelmark M, Hellmark T. *Pseudomonas aeruginosa* in cystic fibrosis: Pyocyanin negative strains are associated with BPI-ANCA and progressive lung disease. *Journal of Cystic Fibrosis*. 2011; 10(4):265–71. <https://doi.org/10.1016/j.jcf.2011.03.004> PMID: 21463973
82. Kim E-J, Wang W, Deckwer W-D, Zeng A-P. Expression of the quorum-sensing regulatory protein LasR is strongly affected by iron and oxygen concentrations in cultures of *Pseudomonas aeruginosa* irrespective of cell density. *Microbiology (Reading, England)*. 2005; 151:1127–38. <https://doi.org/10.1099/mic.0.27566-0> PMID: 15817780
83. O'Toole GA. Microtiter dish biofilm formation assay. *J Vis Exp*. 2011;(47):2437. <https://doi.org/10.3791/2437> PMID: 21307833.
84. Münch R, Hiller K, Barg H, Heldt D, Linz S, Wingender E, et al. PRODORIC: prokaryotic database of gene regulation. *Nucleic Acids Res*. 2003; 31(1):266–9. PMID: 12519998
85. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004; 14(6):1188–90. <https://doi.org/10.1101/gr.849004> PMID: 15173120.