MICROBIOLOGY

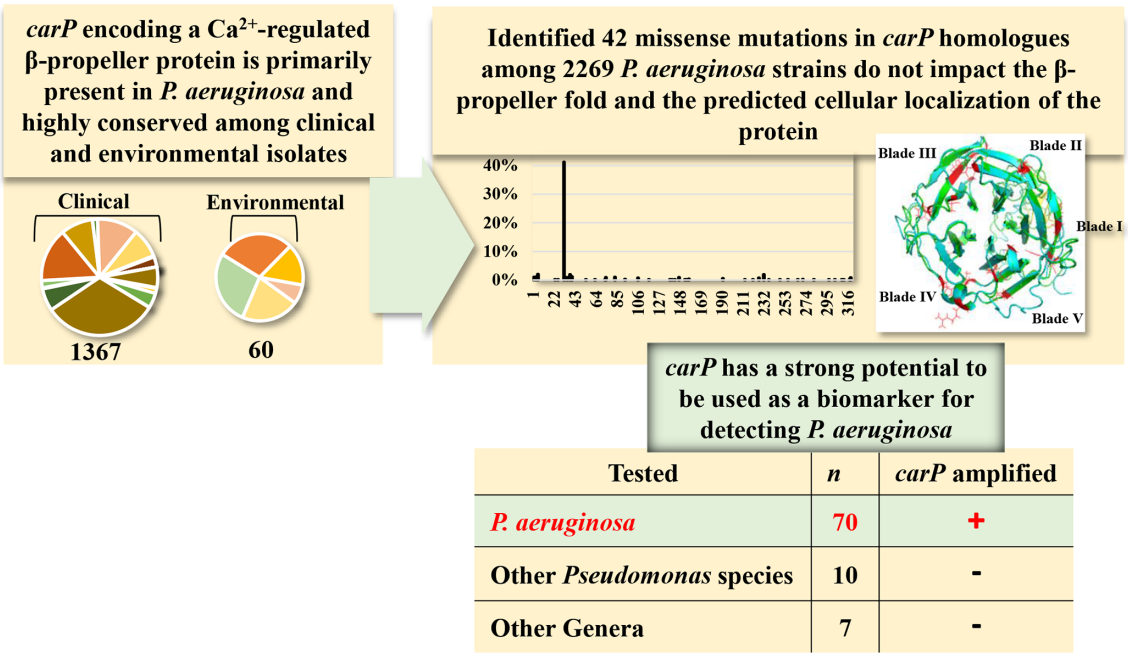
RESEARCH ARTICLE

Mares *et al.*, *Microbiology*

DOI 10.1099/mic.0.001004

# *carP*, encoding a $Ca^{2+}$-regulated putative phytase, is evolutionarily conserved in *Pseudomonas aeruginosa* and has potential as a biomarker

Sergio E. Mares†, Michelle M. King†, Aya Kubo, Anna A. Khanov, Erika I. Lutter, Noha Youssef and Marianna A. Patrauchan*




| Tested | *n* | *carP* amplified |
|---|---|---|
| *P. aeruginosa* | 70 | + |
| Other *Pseudomonas* species | 10 | - |
| Other Genera | 7 | - |

**Graphical abstract**

## Abstract

*Pseudomonas aeruginosa* infects patients with cystic fibrosis, burns, wounds and implants. Previously, our group showed that elevated $Ca^{2+}$ positively regulates the production of several virulence factors in *P. aeruginosa*, such as biofilm formation, production of pyocyanin and secreted proteases. We have identified a $Ca^{2+}$-regulated $\beta$-propeller putative phytase, CarP, which is required for $Ca^{2+}$ tolerance, regulation of the intracellular $Ca^{2+}$ levels, and plays a role in $Ca^{2+}$ regulation of *P. aeruginosa* virulence. Here, we studied the conservation of *carP* sequence and its occurrence in diverse phylogenetic groups of bacteria. *In silico* analysis revealed that *carP* and its two paralogues PA2017 and PA0319 are primarily present in *P. aeruginosa* and belong to the core genome of the species. We identified 155 single nucleotide alterations within *carP*, 42 of which lead to missense mutations with only three that affected the predicted 3D structure of the protein. PCR analyses with *carP*-specific primers detected *P. aeruginosa* specifically in 70 clinical and environmental samples. Sequence comparison demonstrated that *carP* is overall highly conserved in *P. aeruginosa* isolated from diverse environments. Such evolutionary preservation of *carP* illustrates its importance for *P. aeruginosa* adaptations to diverse environments and demonstrates its potential as a biomarker.

## INTRODUCTION

*Pseudomonas aeruginosa* resides in diverse environmental niches, but is mostly known as a human opportunistic pathogen that infects individuals with a variety of underlying medical conditions, including cystic fibrosis (CF) [1], HIV-1 [2] and causes severe corneal and urinary tract infections [3, 4]. In 2017, the Centers for Disease Control (CDC) recognized multi-drug-resistant *P. aeruginosa* as a pathogen of a critical priority [5]. *P. aeruginosa* produces an arsenal of virulence factors [6, 7, 8], and its resistance to antibiotics and host immune protection further enhances its virulence [9, 10, 11]. Current research efforts are focused on *P. aeruginosa* regulatory and signalling networks that enhance the pathogen's virulence and resistance in the host.

Calcium ions ($Ca^{2+}$) serve multiple functions in living organisms and are particularly important for cell signalling. In eukaryotes, $Ca^{2+}$ signalling regulates most essential cellular functions (reviewed in [12]) and therefore, aberrations in $Ca^{2+}$ homeostasis are commonly associated with human diseases, as exemplified by CF, where $Ca^{2+}$ imbalance alters the immune response (reviewed in [13]), which can predispose to bacterial infections. Previously, our group reported that elevated $Ca^{2+}$ increases the production of several virulence factors in *P. aeruginosa* [14, 15, 16]. We have also shown that *P. aeruginosa* tightly regulates intracellular $Ca^{2+}$ homeostasis, which mediates $Ca^{2+}$ regulation of its physiology [17]. We have identified several components of the $Ca^{2+}$ signalling network that can be grouped into two categories: proteins with eukaryotic orthologs ($Ca^{2+}$ transporters [17] and $Ca^{2+}$-binding proteins) and proteins with no eukaryotic orthologues, exemplified by a $Ca^{2+}$-regulated putative $\beta$-propeller phytase, designated as CarP (PA0327) [15].

Sequence analysis of CarP has indicated the presence of three domains: a transmembrane helix domain (TMH), a phytase-like domain, and an SdiA-regulated domain, suggesting its potential role as a phytase that is regulated by a quorum-sensing system [15]. Earlier, we showed that, in addition to being regulated by a $Ca^{2+}$-induced two-component system CarSR, CarP contributes to the regulation of intracellular $Ca^{2+}$ homeostasis and is required for *P. aeruginosa* tolerance to elevated $Ca^{2+}$ [15]. Furthermore, disruption of *carP* reduced $Ca^{2+}$-induced swarming motility and production of pyocyanin, potent virulence factors [15]. Taken together, these data suggest the important role of CarP in $Ca^{2+}$-regulated virulence and adaptations to environments with elevated $Ca^{2+}$, such as those in a human host.

During the course of infection, *P. aeruginosa* acquires numerous mutations. Some mutations cause alterations in gene expression, which lead to enhanced adaptations to the host and thus enable persistent infections [18, 19]. These include activation of genes essential for survival during infections, for example, antibiotic resistance genes, *mexXY, armZ* and *ftsH* [20], or inactivation of genes whose impact on survival declines with progression of infection, for example, *lasR* [21, 19]. Whereas other genes that are also important for the pathogen's survival within a host can be preserved from mutations. These genes have a potential to serve as therapeutic drug targets or as biomarkers.

The early established and commonly used strategies for detecting *P. aeruginosa* in clinics mostly rely on using selective growth media, phenotypic characterization, or amplification and sequencing of *16S rRNA* gene, *oprL* [22], *algD* [23] or *toxA* genes [24]. More recently proposed methods for identifying *P. aeruginosa* include detection of pyocyanin by a silver/gold chip [25], mass spectrometry (MS) of volatile compounds, such as 2-aminoacetophenone, in patient breath samples [26, 27] and detection of specific host proteins produced in response to *P. aeruginosa* infections [28]. While these methodologies are proven useful, they require expensive equipment and processing. Therefore, there is still a need for fast, accurate and inexpensive ways to detect *P. aeruginosa* in clinical samples. This is of particular significance for CF patients, where appearance of this pathogen in the lungs commonly leads to worsening of prognosis [29] and requires timely therapy adjustments.

In this study, we explored the unique nature of *carP* in pseudomonads and, particularly, in *P. aeruginosa*. We performed a comparative sequence analysis of *carP* homologues representing isolates from diverse sources to determine whether *carP* sequence conservation reflects potential adaptation to specific ecological niches. We also tested the potential of *carP* as a biomarker for identifying *P. aeruginosa* in environmental and clinical samples. Overall, the results showed that *carP* is highly conserved, belongs to the core genome of *P. aeruginosa,* and may serve as a biomarker for detecting this pathogen in diverse settings.

## METHODS

### Bacterial strains and growth conditions

Ten CF patient sputa samples were collected at the OU Children's Hospital in Oklahoma City, OK, USA and plated onto *Pseudomonas* isolation agar (PIA, BD Biosciences). The samples represented ten patients, from 6 to 55 years old

Received 06 September 2020; Accepted 24 November 2020; Published 09 December 2020
Author affiliations: [1]Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74078, USA.
*Correspondence: Marianna A. Patrauchan, m.patrauchan@okstate.edu
Keywords: biomarker; cystic fibrosis and ocular clinical isolates; calcium; putative phytase; sequence conservation.
Abbreviations: CDC, Centers for Disease Control and Prevention; CF, cystic fibrosis; NCBI, The National Center for Biotechnology Information; TMH, transmembrane helix.
†These authors contributed equally to this work
Three supplementary figures and one supplementary table are available with the online version of this article.

**Table 1.** PCR amplification of *carP* in diverse bacterial species

| Strains | Isolation source | PCR *carP* | PCR *16S rRNA* | Source/reference |
|---|---|---|---|---|
| *P. aeruginosa* PAO1 | Wound isolate | + | + | [55] |
| *P. aeruginosa* PA14 | Human isolate | + | + | [56] |
| *P. aeruginosa* FRD1 | CF isolate | + | + | [57] |
| *P. aeruginosa* OSDH 1 | Blood | + | + | Oklahoma State Department of Health |
| *P. aeruginosa* OSDH 2 | Trachea | + | + | Oklahoma State Department of Health |
| *P. aeruginosa* OSDH 3 | Blood | + | + | Oklahoma State Department of Health |
| *P. aeruginosa* OSDH 4 | Wound | + | + | Oklahoma State Department of Health |
| *P. aeruginosa* | Hot spring | + | + | Fathepure Lab |
| Pooled *Pseudomonas strains* | CF Sputum Sample 5814A | + | + | Lutter Lab |
| Pooled *Pseudomonas strains* | CF Sputum Sample 102314A | + | + | Lutter Lab |
| Pooled *Pseudomonas strains* | CF Sputum Sample 2515D | + | + | Lutter Lab |
| Pooled *Pseudomonas strains* | CF Sputum Sample 6514D | + | + | Lutter Lab |
| Pooled *Pseudomonas strains* | CF Sputum Sample 61214C | + | + | Lutter Lab |
| Pooled *Pseudomonas strains* | CF Sputum Sample 6514C | + | + | Lutter Lab |
| 96 *Pseudomonas strains* | CF Sputum Sample 31314A | + | + | Lutter Lab |
| 96 *Pseudomonas strains* | CF Sputum Sample 22014D | + | + | Lutter Lab |
| 96 *Pseudomonas strains* | CF Sputum Sample 22714D | + | + | Lutter Lab |
| 96 *Pseudomonas strains* | CF Sputum Sample 22014B | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014B_A1 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014B_B1 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014B_H12 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014B_G12 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014B_F12 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014B_E12 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014B_D12 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014B_C12 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014B_H11 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014B_G11 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014B_F11 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014B_H10 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014D_A1 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014D_B1 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014D_A2 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014D_A3 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014D_B4 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014D_B5 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014D_C1 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014D_C3 | + | + | Lutter Lab |

Continued

**Table 1.** Continued

| Strains | Isolation source | PCR *carP* | PCR *16S rRNA* | Source/reference |
|---|---|---|---|---|
| *P. aeruginosa* | CF Sputum Sample 22014D_C4 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22014D_E1 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 31314A_A1 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 31314A_B1 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 31314A_A2 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 31314A_A3 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 31314A_A4 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 31314A_A5 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 31314A_B3 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 3 314A_B6 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 31314A_E6 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22714D_A1 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22714D_D1 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22714D_B2 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22714D_C2 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22714D_E3 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22714D_F3 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22714D_C3 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22714D_D1 | + | + | Lutter Lab |
| *P. aeruginosa* | CF Sputum Sample 22714D_D2 | + | + | Lutter Lab |
| *P. aeruginosa* | Ocular infections_E1 | + | + | Callegan Lab |
| *P. aeruginosa* | Ocular infections_E2 | + | + | Callegan Lab |
| *P. aeruginosa* | Ocular infections_H3 | + | + | Callegan Lab |
| *P. aeruginosa* | Ocular infections_C2 | + | + | Callegan Lab |
| *P. aeruginosa* | Ocular infections_F4 | + | + | Callegan Lab |
| *P. aeruginosa* | Ocular infections_H5 | + | + | Callegan Lab |
| *P. aeruginosa* | Ocular infections_C4 | + | + | Callegan Lab |
| *P. aeruginosa* | Ocular infections_C1 | + | + | Callegan Lab |
| *P. aeruginosa* | Ocular infections_C3 | + | + | Callegan Lab |
| *P. aeruginosa* | Ocular infections_F3 | + | + | Callegan Lab |
| *P. aeruginosa* | Ocular infections_C11 | + | + | Callegan Lab |
| *P. aeruginosa* | Ocular infections_G10 | + | + | Callegan Lab |
| *P. fluorescens Pf-5* | Soil | – | + | [58] |
| *P. protegens CHAO* | Soil | – | + | [59] |
| *P. fluorescens SBW25* | Soil, Water, Plant | – | + | [59] |
| *P. chlororaphis 30–84* | Soil | – | + | [59] |
| *P. veronii 24* | Stream | – | + | Mavrodi Lab |

Continued

**Table 1.** Continued

| Strains | Isolation source | PCR *carP* | PCR *16S rRNA* | Source/reference |
|---|---|---|---|---|
| *P. aridius R1-43-08* | Rhizosphere | – | + | [60] |
| *P. mosselii KM2404* | Clinical Strain | – | + | [61] |
| *P. synxantha 2–79* | Rhizosphere | – | + | [62] |
| *P. brassicacearum* | Soil | – | + | [59] |
| *P. monteilii* | Lab Sink | – | + | Patrauchan Lab |
| *E. coli* DH5α | Lab Strain | – | + | Patrauchan Lab |
| *S. aureus* strain NRS70 | Lab Strain | – | + | [63] |
| *M. luteus* | Lab Strain | – | + | OSU Teaching Lab |
| *K. oxtoca* | Lab Strain | – | + | OSU Teaching Lab |
| *B. aryabhattai* | Lab Strain | – | + | [63] |
| *Arthrobacteria* sp. | Rumen | – | + | Fathepure Lab |
| *Alcaligenes* sp. *3K* | Nigeria | – | + | Fathepure Lab |

*The identity of the isolates was verified by sequencing their amplified *16S rRNA* gene.

(Table 1). For the first group including four sputum samples, 96 randomly selected individual isolates for each sample were stored frozen in 10% (w/v) skim milk in 96-well blocks. The identity of at least ten random isolates from each plate was verified by sequencing their *16S rRNA* gene. To study *carP* sequence conservation during infection, the 96 isolates from each plate were grown and pooled as described below. For the second group of six sputum samples, individual strains were not isolated, and instead, all colonies grown on PIA were pooled for the following *carP* sequencing and analyses.

A total of 70 individual *Pseudomonas* strains were tested for the presence of *carP* (Table 1). They included 40 sequence-verified *P. aeruginosa* CF isolates randomly selected from the 96-well blocks described above; 12 *P. aeruginosa* isolates from ocular infections obtained from Dr Michelle Callegan, University of Oklahoma Health Sciences Center; four *P. aeruginosa* isolates from blood, trachea and wound provided by Oklahoma State Department of Health; 11 *Pseudomonas* environmental isolates mostly shared by Drs Dmitri and Olga Mavrodi, University of Southern Mississippi; three common laboratory strains; and one *P. monteilii* strain isolated from the laboratory sink. In addition, we included seven non-*Pseudomonas* species. The identity of strains in question was determined or verified by *16S rRNA* gene sequencing. The universal *16S rRNA* primers 27F/1492R [30] were obtained from Integrated DNA Technologies (Table 2). Prior to PCR analyses, bacterial strains were either struck on Luria–Bertani (LB, VWR) plates or grown in LB in 96-well plates for 16–24 h at 37 °C.

The deletion mutant, Δ*carP,* was generated using Gibson cloning [31]. In brief, 500 bp regions flanking *carP* were amplified using the primers *carP*upF, *carP*upR, *carP*downF, *carP*downR (Table 2) and cloned into pEXG2 by using the HiFi DNA assembly kit (NEB). The generated plasmid p*carP*DM was then electroporated into *E. coli* SM10 for biparental mating with PAO1. The successful transformants were selected on *Pseudomonas* Isolation Agar (PIA) with 30 μg ml$^{-1}$ gentamicin. To facilitate homologous recombination, the cultures were grown in liquid LB for 6 h followed by SacB-based counter selection using LB agar supplemented with 5% sucrose. The loss of the gene in the sucrose-sensitive and gentamicin-resistant transformants was verified by PCR using gene-specific primers *carP*upF and *carP*downR (Table 2).

## Gene-expression analysis by RT-PCR

Wild-type *P. aeruginosa* PAO1 and Δ*carP* mutant were grown in Biofilm Mineral Media (BMM) [14] until mid-log phase (12 h; $OD_{600}$=0.18±0.03). Total RNA was isolated, and cDNA was synthesized as previously described [16]. Briefly, total RNA was isolated from *P. aeruginosa* mid-log cultures grown in BMM grown without $Ca^{2+}$ or with 5 mM $Ca^{2+}$ by using RNeasy Protect Bacteria Mini kit (Qiagen). The quality of the purified RNA was assessed by 1.2% agarose gel electrophoresis. The absence of genomic DNA in total RNA samples was confirmed by using DreamTaq Green PCR Master Mix (Thermo Fisher Scientific) and a primer set *16S*-533-F /*16S*-1100-R for the amplification of the *16S rRNA* gene (Table 2). For reverse transcription-PCR (RT-PCR) reactions, 1 μl of the cDNA was used with *carP*-, PA0319- and PA2017-specific primer sets F_*carP*_+21/New_*carP*_R, PA0319_123_F/PA0319_990_R and PA2017_1_F/PA2017_476_R, respectively (Table 2). The *16S rRNA* gene was amplified as an internal control. Amplification was performed using Dream Taq Green PCR Master Mix following the manufacturer's instructions and

**Table 2.** Primers used in this study

| Primer name | Sequence | Source |
|---|---|---|
| F_*carP* | AGAGAGCATATGATGACTATCCACGCC | This study |
| R_*carP* | AGAGAGGGATCCTCAGGGCTGCGC | This study |
| - 326_XhoI 5' | TAAGCACTCGAGGCTGAATTCCATGCA | This study |
| *carP*up_F | TGCGCACCCGTGGAAATTAATTAAGGTACCGAATTC CGGCAGGCCTTCCATGTGTTC | This study |
| *carP*up_R | GACGCTCCTGAGCATAGAGCCTGTACGCTG | This study |
| *carP*down_F | GCTCTATGCT CAGGAGCGTCCCCTGGCC | This study |
| *carP*down_R | TTATACGAGCCGGAAGCATAAATGTAAAGC AAGCTTAGCGTCTGTCGGGCGACAC | This study |
| F_*carP*_+21 | CCCGAGCCGTTTTCCATGTC | This study |
| New_*carP*_R | TCAGGGCTGCGCCGGCTCGTC | This study |
| PA0319_123_F | CGTGGTGCTGACCATCGTC | This study |
| PA0319_990_R | TTAATCGGCGTCGCTGGTGGC | This study |
| PA2017_1_F | ATGAGTCGGCTGATGACCTACAGG | This study |
| PA2017_476_R | AACTTCCGGGCGTTGCGCG | This study |
| *16S*-533-F | GTGCCAGCAGCCGCGGTAA | [64] |
| *16S*-1100-R | AGGGTTGCGCTCGTTG | [65] |
| *16S* 27F | AGAGTTTGATCCTGGCTCAG | [30] |
| *16S* 1492R | ACGGCTACCTTGTTACGACTT | [30] |

the following protocol for annealing: 56 °C, 35 cycle for *carP*; 63 °C, 30 cycle for PA0319; 72 °C, 35 cycle for PA2017; 56 °C, 16 cycle for *16S rRNA* gene. The PCR products were subjected to electrophoresis in 1.0% agarose gel, stained with ethidium bromide, and the signals were quantified using ImageJ software (NIH, USA). RT-PCR was repeated three times for consistency. A representative gel image is presented. Densitometry analysis was performed based on three independent experiments using ImageJ software [32]. The intensities of *carP,* PA0319 and PA2017 bands were averaged and normalized by the intensity of the *16S rRNA* gene.

## Sequence analyses and predictions

At the time of the analyses, the *Pseudomonas* database v. 17.2 (www.*pseudomonas.com*) [33] consisted of 3348 complete and incomplete genomes of 426 *Pseudomonas* species. To test the sequence conservation of *carP* among *Pseudomonas* species, we used BLASTN and BLASTP [34] and considered homologues that share at least 20% sequence identity over the full length of the gene with the E-value of 10 as a cutoff. Additional sequence alignments were performed using ClustalW [35]. The taxonomy of strains carrying homologues was either verified by using the software GTDB-Tk [36] or by calculating their average nucleotide identity values (ANI) in Kbase [37]. To identify integrative conjugative elements (ICE), we used ICEfinder algorithms run through the web version of ICEberg 2.0 [38] To identify the residues naturally mutated in *carP* in diverse isolates and find the regions of high conservation, the multi-sequence alignment software MEGA 7.0 [39] was used. The isolation source for each of these sequences was extracted using an in-house python script programmed by Eric King [40]. To identify and visualize single nucleotide mutations within *carP* homologues in verified *P. aeruginosa* sequences, Unipro UGENE software was used [41]. The nucleotide sequences were translated by using MEGA, and mutations in amino acids were identified and visualized by Unipro UGENE. Sequences were clustered using CD-HIT suite [42] at 99.5% similarity cutoff. Aligned sequences were used to construct maximum-likelihood phylogenetic trees based on JTT matrix model [43] in MEGA 7.0 [39]. The secondary structures of CarP and its homologues were predicted by CFSSP [44]. Protein three-dimensional (3D) structure was predicted using I-TASSER [45] and SWISS-MODEL. I-TASSER predicts 3D structure using Protein Data Bank templates and the multiple threading approach LOMETS. The predicted structures were visualized using PyMOL [46] (version 1.8.6.0; Schrödinger, LLC). The overall hydrophobicity and the transmembrane tendency of TMH domain in CarP and its *in silico* mutated version were predicted by using ExPASy Bioinformatics Resource Portal. Selection pressure was estimated for each codon using the HyPhy method [47] implemented in MEGA 7.0 [39]. In addition,

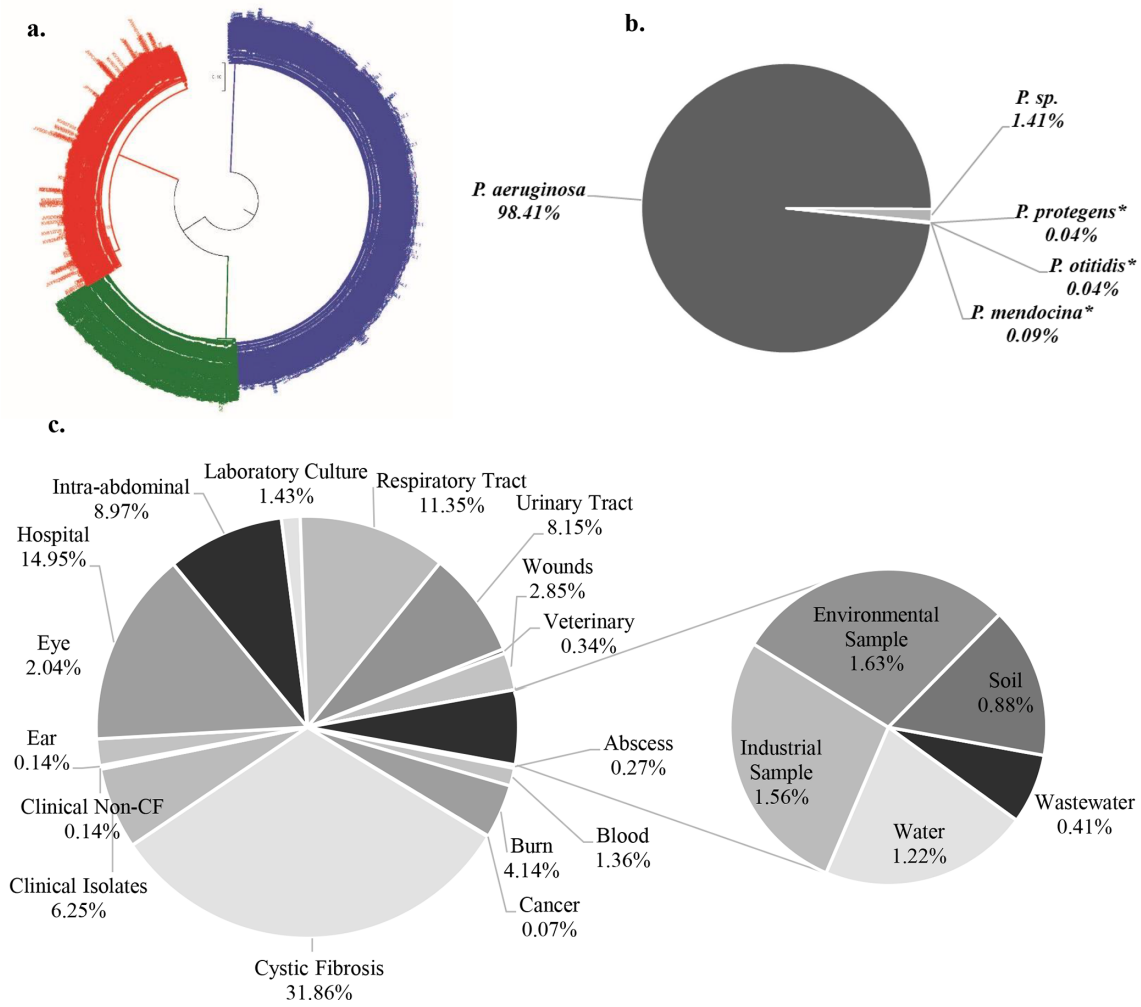



**Fig. 1.** Phylogenetic analysis of sequences obtained from *in silico* analysis. (a) The 7665 sequences identified as significant hits to *carP* using BLASTN were aligned using MAFFT and the alignment was used to construct a maximum-likelihood phylogenetic tree using FastTree. The tree is midpoint rooted. The sequences clearly clustered into three groups; group 1 (shown in red) with 95.34–100% sequence identity to *carp*, group 2 (shown in green) with 98.7–100% sequence identity to PA2017 and group 3 (shown in blue) with 98–100% sequence identity to PA0319. (b) Distribution of *carP* homologues in *Pseudomonads*. Overall, 2269 out of 3348 *Pseudomonas* genomes containing *carP* homologues that share at least 96% sequence identity. *P.* sp represents *Pseudomonads* that have not been identified to species. The ANI analyses revealed that all *P. sp.* isolates belong to *P. aeruginosa*. *P. protegens* (NZ_JVPC01000103), *P. otitidis* (NZ_JGYF01000074), and *P. mendocina* (NZ_JVSG01000050, NZ_JVWH01000013) were later reclassified as *P. aeruginosa*. (c) Distribution of *P. aeruginosa* strains harbouring *carP* homologues according to their isolation source. A total of 1427 isolates with the available information on isolation sources were included into the analysis. Isolate-search script was used to extract the information about the isolation source for each sequence [40].

an analysis of the rates of non-synonymous versus synonymous mutations was conducted using codon-based Z-test of selection in MEGA 7.0 [39]. The entire full-length *carP* gene was compared to truncated regions representing the transmembrane helix (TMH) (bases 68–129), and the SdiA-regulated (bases 187–942) domains to identify portions of the gene that are potentially experiencing selection.

## Detecting *carP* by PCR

To study *carP* sequence variability during infection, we used sputum samples from ten CF patients of different ages (Table 1). For each patient, a pooled sample of clinical isolates was prepared for DNA isolation, *carP* amplification and sequencing. For each sample of the first group (2.1), 96 individual *Pseudomonas* isolates were grown in LB in a 96-well plate for 16–24 h. After measuring $OD_{600}$, the cultures were pooled for DNA extraction. For their equal representation, their volumes were proportional to their $OD_{600}$. Each sample of the second group (2.1) was struck on *Pseudomonas* selective PIA plates, and for each, all grown isolates were pooled and prepared for DNA extraction. DNA was extracted by using Wizard Genomic

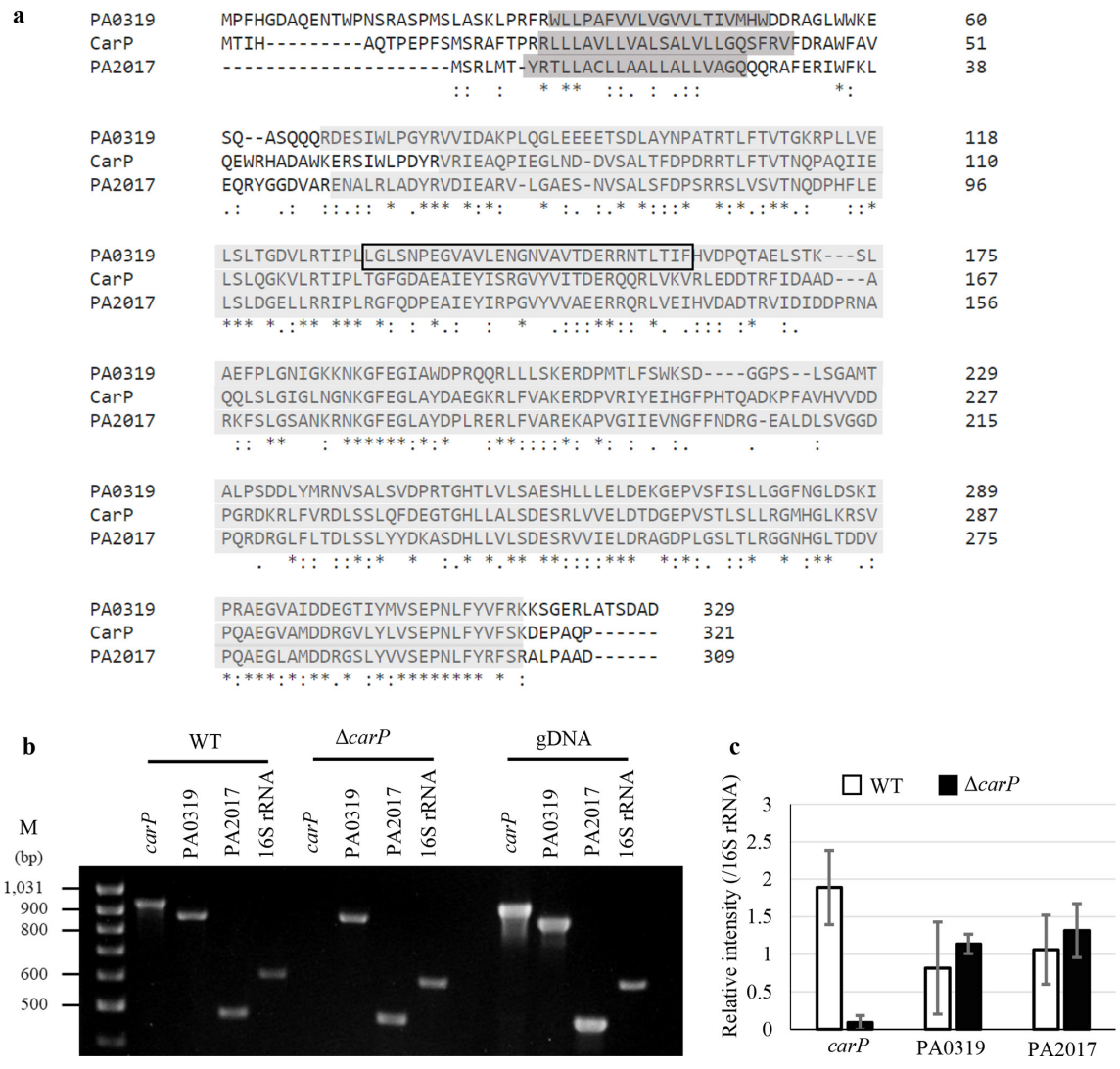



**Fig. 2.** Comparison of CarP and homologues in *P. aeruginosa*. (a) Alignment of CarP, PA2017 and PA0319 sequences using ClustalOmega [66]. TMH is highlighted in light grey, phytase-like domain in dark grey and the NHL repeat domain is outlined in black. Asterisks indicate identical residues and the dots represent similar residues. (b) RT-PCR analysis of *carP*, PA0319, PA2017 and *16S rRNA* gene in WT PAO1 and Δ*carP* strains. *16S rRNA* gene was used as a control. M, MassRuler DNA Ladder (Thermo Scientific). RT-PCR was repeated three times, each with consistent results. A representative gel is presented. (c) Densitometry analysis was performed based on three independent experiments by using ImageJ software [32]. The intensities of *carP*, PA0319 and PA2017 bands were averaged and normalized by the intensity of the *16S rRNA* gene.

DNA Purification Kit (Promega, MD) according to the manufacturer's instruction. The *carP* primers (Table 2) were designed and their specificity was tested by using BLASTN against the NCBI nr and *Pseudomonas* databases *in silico* and verified by using the deletion *carP* mutant. Physical parameters and quality of the primers were analysed by using the IDT OligoAnalyzer. All primers were obtained from IDT.

To verify the presence of *carP* in the individual isolates, 70 individual *Pseudomonas* strains were selected and grown on LB agar plates. Colony-based PCR amplification [48] with specific primers for *16S rRNA* and *carP* genes was conducted. The *16S rRNA* amplicons were sequenced at the OSU DNA Sequencing Core Facility.

PCR amplification was conducted in 25 µl reaction mixtures containing: 15.5 µl of nuclease free $H_2O$, 5.0 µl of 5 x GC buffer (Thermo Fischer Scientific, F519L), 0.5 µl of 10 mM dTNPs, 0.75 µl DMSO, 0.5 µl Phusion polymerase, 1.25 µl of each forward and reverse primers. PCR amplification was initiated by denaturation for 30 s at 98 °C, followed by 30 cycles of denaturation at 98 °C for 30 s, annealing at 63 °C for 30 s, and elongation at 72 °C for 10.5 min. For *16S rRNA* amplification, the PCR reaction followed the same protocol, except for the melting temperature of 56 °C. All

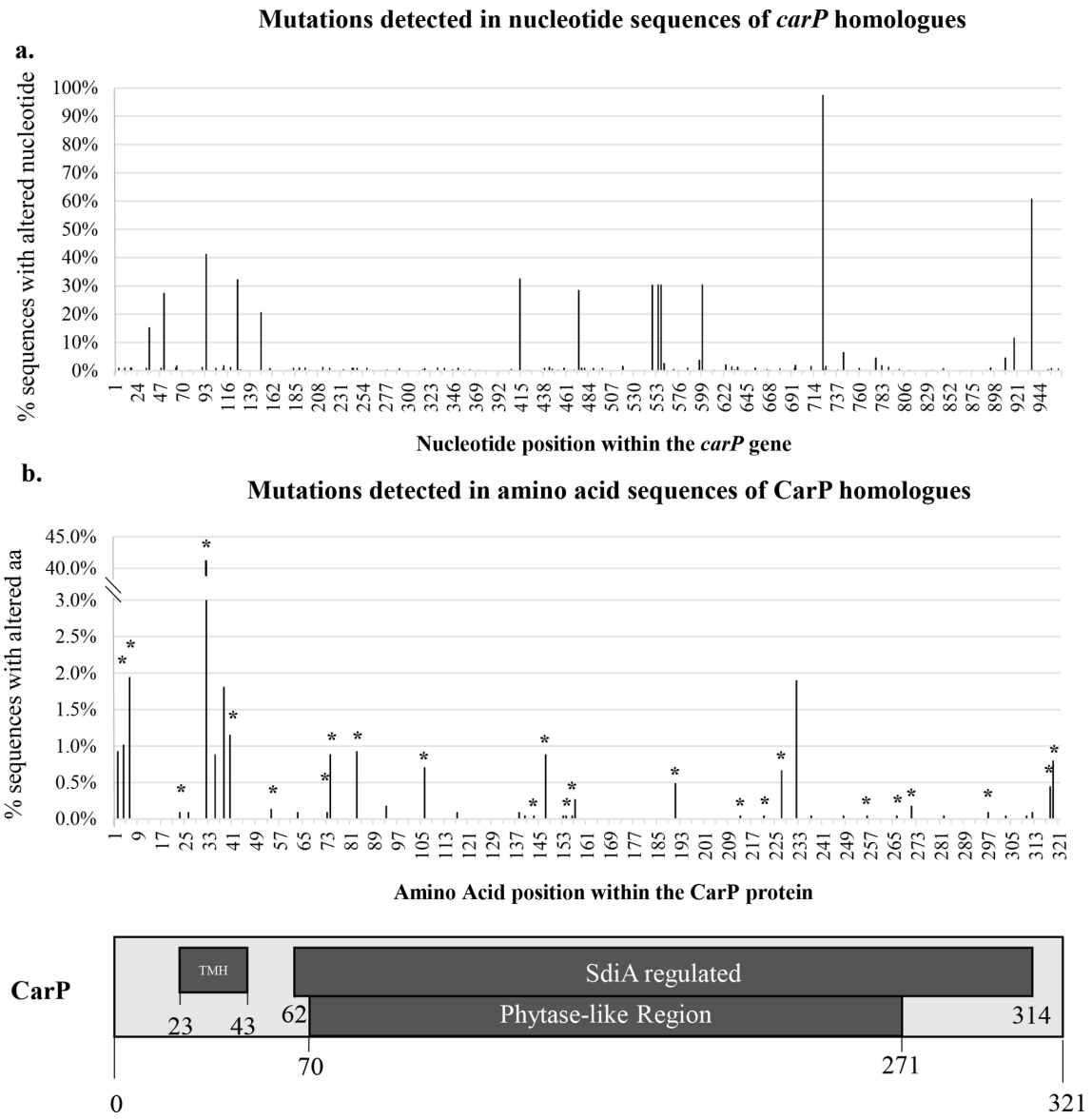



**Fig. 3.** Mutations identified in 2269 *carP* homologues. (a) The grid profiles of the mutations identified in the nucleotide sequences of *carP* homologues graphed as a percent of sequences with altered nucleotide at each position among 2269 sequence in comparison to the *carP* sequence in PAO1. (b) The 2269 nucleotide sequences were translated to amino acid sequences and graphed in a grid profile showing the missense mutations. White gaps indicate 100% conservation in all sequences. * indicate the mutations with a normalized d*N*-d*S* values higher than 1, indicating a positive selection (Table S1).

PCR reactions were carried out using Applied Biosystems 2720 Thermal cycler.

## RESULTS AND DISCUSSION

### *carP* is present primarily in *P. aeruginosa* genomes

Surveying the non-redundant NCBI database for *carP* homologues revealed that the gene homologues are only present in the genomes of *Pseudomonas* genus and no homologous genes were identified in other genera. We have also performed BLASTP analyses, and a few significant hits were returned as non-*P. aeruginosa*. However, upon further analysis using both the genome classification software GTDB-TK and ANI analysis, we verified that all these hits were misclassified and in fact belong to the species *P. aeruginosa*. To study the conservation of *carP* sequence amongst *Pseudomonas* species, we performed BLASTN analysis using the *Pseudomonas* database with, at the time of the analyses, a total of 3348 complete and incomplete genomes representing 426 identified species and 2197 strains of *P. aeruginosa*. The BLASTN alignments yielded 7665 significant hits with sequence identity ranging from 54–100%. Considering the presence of two *carP*-homologous genes in PAO1 genome (PA2017 and PA0319), we binned

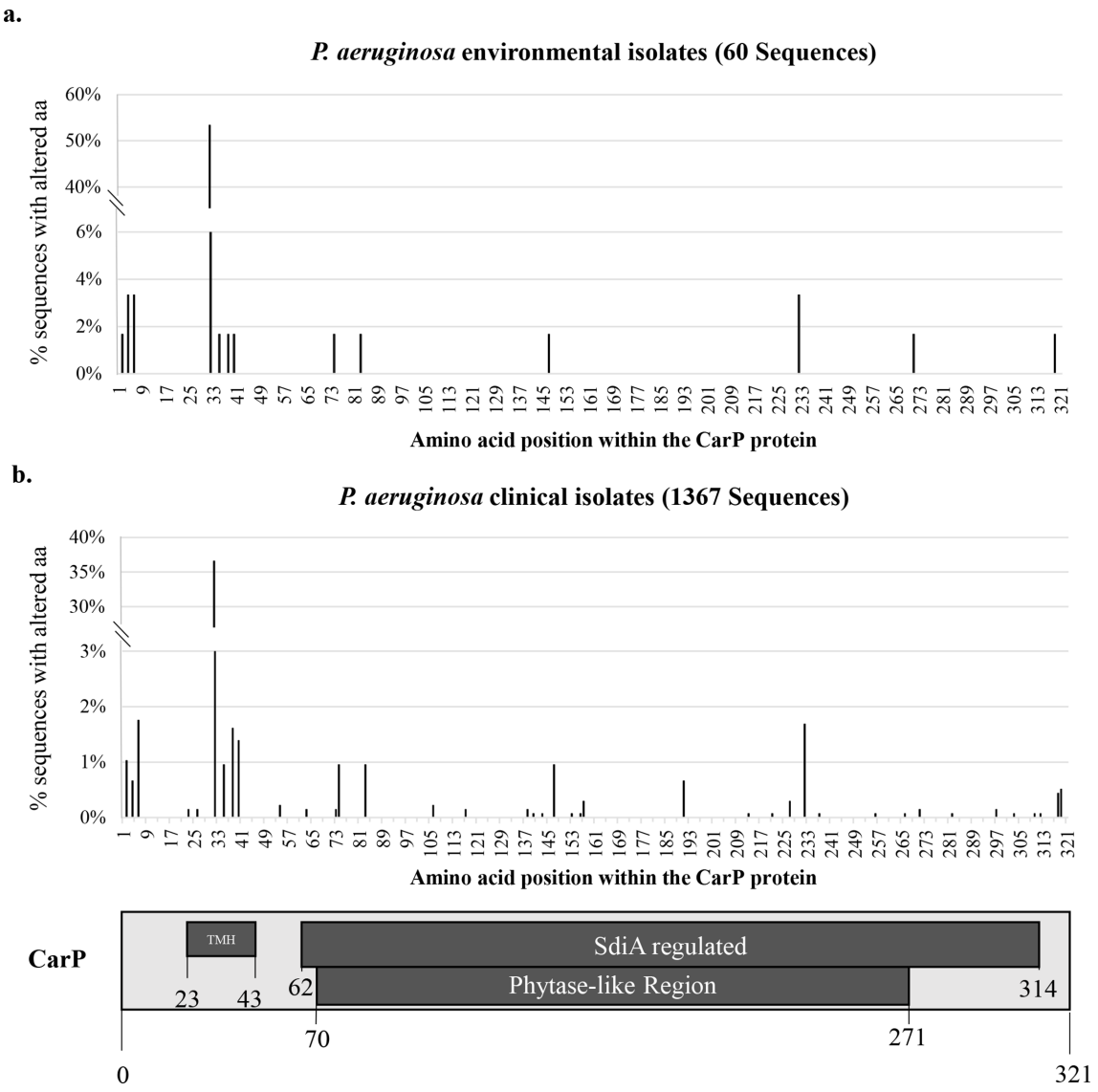



**Fig. 4.** Protein sequence alterations in CarP homologues detected in 60 environmental and 1367 clinical isolates of *P. aeruginosa*. (a) Grid profiles of mutations identified in 60 environmental isolates graphed as a percent of sequences with altered nucleotide at each position among 60 sequences in comparison to the CarP sequence in PAO1. (b) Grid profiles of mutations identified in 1367 clinical isolates graphed as a percent of sequences with altered nucleotide at each position among 1367 sequence in comparison to the CarP sequence in PAO1. White gaps indicate 100% conservation in all sequences.

the sequences into three groups (Fig. 1a). The first group included 2269 sequences that shared 95–100% sequence identity with *carP*, but below 70% identity with PA2017 and PA0319. The majority of the sequences (*n*=2218) in this group belong to *P. aeruginosa* (Fig. 1b). The three other species, *P. protegens* (NZ_JVPC01000103; one sequences), *P. otitidis* (NZ_JGYF01000074; one sequence) and *P. mendocina* (NZ_JVSG01000050, NZ_JVWH01000013; two sequences) have been reclassified as *P. aeruginosa*. Forty-eight strains in this group were not identified at the species level. We verified the taxonomy of these 48 strains by calculating their average nucleotide identity values (ANI) using the software GTDB-Tk [36]. They all showed ANI values above 98% (98.04–99.41%) to the reference *P. aeruginosa* strain NCTC10332 (GenBank assembly accession number GCF_001457615.1), supporting their classification as *P. aeruginosa*. Thus, we concluded that the first group includes only *P. aeruginosa* species.

The second group of 1394 sequences shared 98–100% identity with PA2017. This group is represented only by *P. aeruginosa* strains. Because of the higher similarity to PA2017, we attributed them to be homologous primarily to PA2017. This group is 71.3–71.7% identical to *carP*, and 58.2–59% identical to PA0319. The third group contains 4002 sequences that are 98–100% identical to PA0319. This group

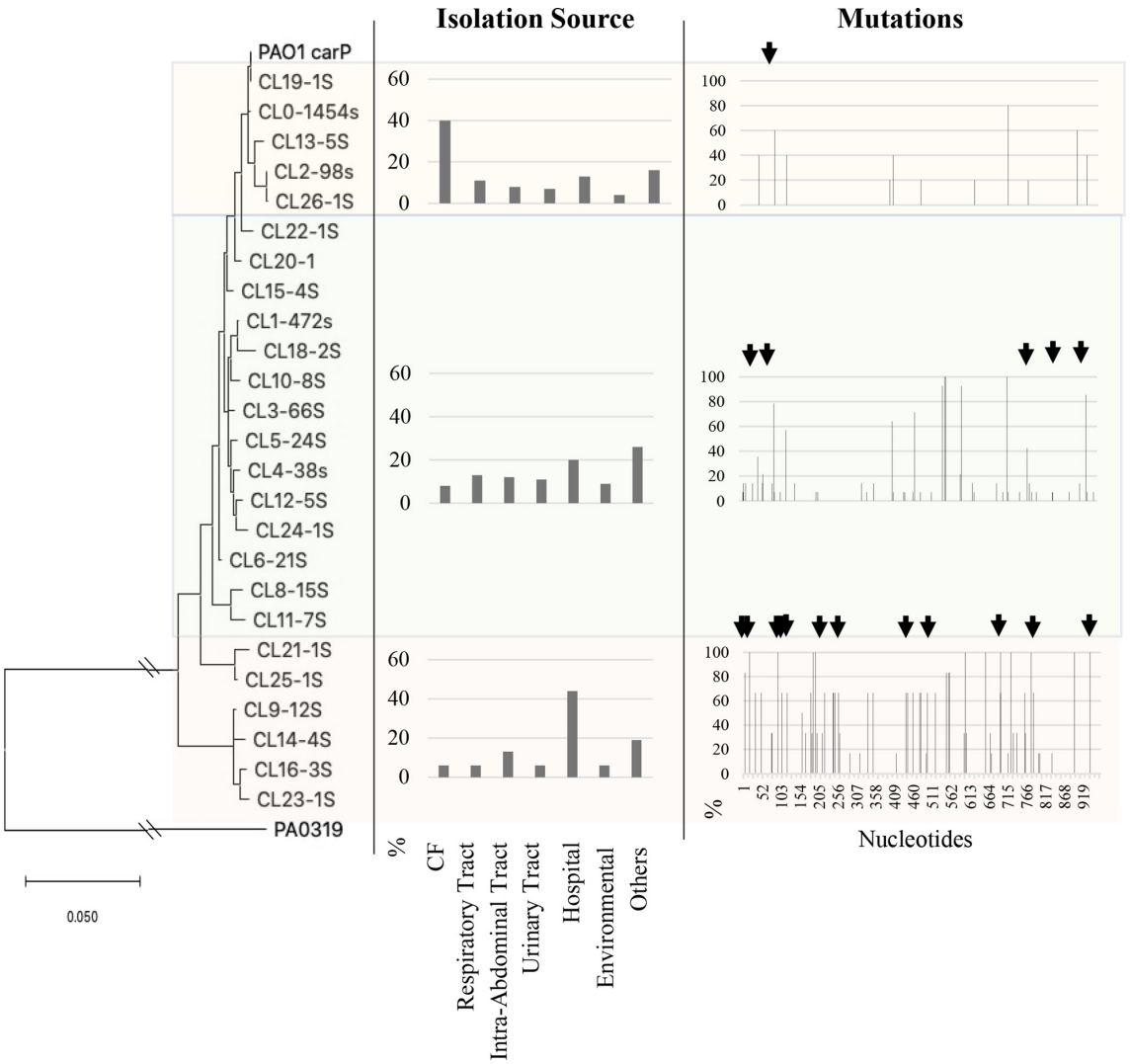



**Fig. 5.** Correlative analyses of phylogenetic relationship, isolation sources and frequency of mutations in *carP* homologues. The 2248 sequences of *carP* homologues and *carP* were clustered using CD-Suite [42]. One representative from each cluster was used to build a phylogenetic tree in MEGA [39]. The clustering together with distribution of the isolation sources formed three apparent groups depicted using the shaded boxes. Top group represents 1559 sequences, middle group represents 667 sequences, bottom group represents 22 sequences. The isolation sources shown are those that represent at least 8% of the sequences shown in Fig. 1c. All other sources are represented as others. Mutation profiles are depicted for each group. The arrows indicate missense mutations.

is 57.6–58.4% identical to *carP,* and 55.8–56.5% identical to PA2017 and is comprised of all *P. aeruginosa* strains.

PA2017 and PA0319 are 64 and 53% identical to *carP,* respectively. They share predicted 3D structure and the presence of transmembrane helix, putative phytase and SdiA-regulated domains (Fig. 2a). PA0319 has an additional predicted NHL repeat domain (named after ncl-1, HT2A and lin-41) [49], that the other two proteins do not possess. The GC content of each of these genes is 65–67%, which is similar to the average GC content of *P. aeruginosa* genome (66.6%). No sequence signatures of mobile elements within the immediate genomic environment were identified, which was supported by the genome-wide analyses using ICEfinder and agreed with the previous report [50]. Taken together, these observations suggest that all three genes are native to *P. aeruginosa*. Further, these genes share the *carP* conservation pattern, and are present in all the sequenced *P. aeruginosa* genomes, indicating they belong to the core genome of *P. aeruginosa*.

Based on the sequence similarities, we hypothesized that PA2017 and PA0319 may share a CarP function and serve as a functional backup and that, if *carP* is deleted, their expression may increase to functionally complement the lack of CarP. We observed the precedent for the *P. aeruginosa* $\beta$-carbonic anhydrases [51]. To test this, we performed RT-PCR of the three genes in wild-type PAO1 and a Δ*carP* strain (Fig. 2b, c). However, the RT-PCR results (Fig. 2b, c) showed that all three homologues appeared to be transcribed at similar levels in PAO1, and the expression of PA2017 and PA0319 remained

**Table 3.** Conservation of *carP* sequence among 2269 sequenced *P. aeruginosa* retrieved from *Pseudomonas.com* database*

| Isolation source | No. of seq. | Regions | | | Full gene length |
|---|---|---|---|---|---|
| | | TMH | SdiA regulated | Phytase-like | |
| Abscess | 4 | 100.0% | 100.0% | 100.0% | 99.9% |
| Blood | 20 | 97.1% | 99.98% | 99.98% | 99.8% |
| Burn | 61 | 96.4% | 99.9% | 99.9% | 99.7% |
| Cancer | 1 | 100.0% | 100.0% | 100.0% | 100.0% |
| CF | 469 | 98.6% | 100.0% | 100.0% | 99.9% |
| Clinical isolates | 92 | 96.0% | 100.0% | 100.0% | 99.7% |
| Clinical non-CF | 2 | 95.2% | 100.0% | 100.0% | 99.7% |
| Ear infection | 2 | 90.0% | 98.9% | 98.7% | 98.1% |
| Eye infection | 30 | 96.8% | 100.0% | 100.0% | 99.8% |
| Respiratory tract | 167 | 96.9% | 100.0% | 100.0% | 99.7% |
| Urinary tract | 120 | 96.9% | 100.0% | 100.0% | 99.8% |
| Intra-abdominal | 132 | 96.6% | 99.9% | 99.9% | 99.7% |
| Wounds | 42 | 96.94% | 100.0% | 100.0% | 99.7% |
| Veterinary | 5 | 97.3% | 100.0% | 100.0% | 99.8% |
| Hospital | 220 | 96.2% | 99.9% | 99.9% | 99.6% |
| Laboratory culture | 21 | 96.0% | 100.0% | 100.0% | 99.8% |
| Environmental | 24 | 97% | 100.0% | 100.0% | 99.8% |
| Water | 18 | 97.8% | 100.0% | 100.0% | 99.6% |
| Wastewater | 5 | 95.0% | 100.0% | 100.0% | 99.8% |
| Wastewater | 1 | 95.0% | 100.0% | 100.0% | 99.7% |
| Industrial | 23 | 98.0% | 100.0% | 100.0% | 99.9% |
| Soil | 13 | 97.6% | 100.0% | 100.0% | 99.8% |
| Abscess | 3 | 100.0% | 100.0% | 100.0% | 100.0% |
| N/A | 795 | 96.6% | 99.9% | 99.9% | 99.7% |
| N/A | 2 | 96.7% | 100.0% | 100.0% | 99.8% |

*Percentages represent the nucleotide sequence identity within the corresponding region of the gene in comparison to PAO1 *carP*.

unchanged in the absence of *carP*. Further, according to our previous microarray data, in contrast to *carP,* the expression of PA0319 and PA2017 is not regulated by $Ca^{2+}$, nor it is regulated by the $Ca^{2+}$-induced two component system CarSR [15], suggesting that PA2017 and PA0319 are not a part of the $Ca^{2+}$ response in *P. aeruginosa* and may serve different roles in the cell. In support, all three genes are embedded into different genomic environments as single open reading frames and, according to the genomic data repository, GEO [52], are regulated differently under different growth conditions, with the exception of the three following conditions: during stationary growth phase in artificial medium mimicking CF lung sputum (GDS4250), *in vivo* growth in CF sputum (GDS2869), and in response to airway epithelia (GDS2502). Taken together, this indicates that *carP* and its two paralogues are likely involved in different cellular processes, differentially regulated, but all are important during infection.

## *carP* is conserved in a wide range of *P. aeruginosa* clinical and environmental isolates

Considering the role of CarP in *P. aeruginosa* virulence, we hypothesized that its sequence may be preserved and therefore more conserved among clinical isolates of the species. To test this hypothesis, we first retrieved the information on the isolation sources for each strain carrying a *carP* homologue in group 1 (described above). From the 2269 *P. aeruginosa* strains combined in this group, only 1427 had the information

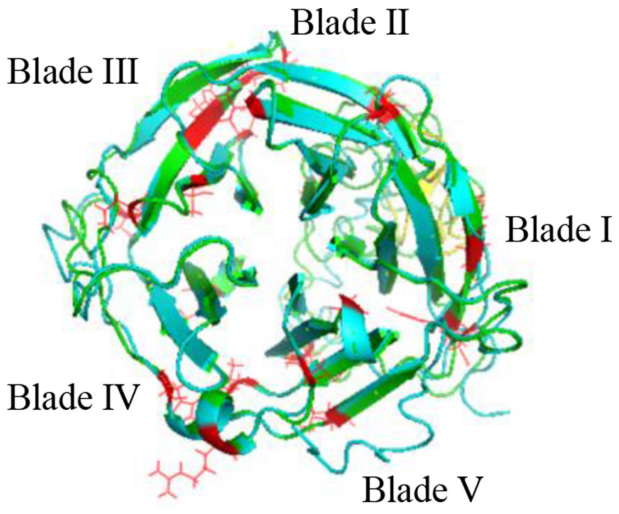



**Fig. 6.** Impact of the detected missense mutations on the predicted 3D structure of CarP. Forty mutations identified in the phytase-like domain during analysis of 2269 *P. aeruginosa* sequenced genomes were introduced *in silico* into CarP sequence and the mutated sequence was submitted to I-Tasser [67]. The predicted structure of the mutated protein is shown in cyan, the wild-type CarP is in green, and the mutations within the sequence are highlighted in red. TMH is highlighted in yellow.

describing their isolation source. The majority of the strains, 93.2%, were isolated from clinical sources (Fig. 1c) with the largest group, 31.9%, isolated from CF patients. The other clinical isolates were obtained from burn wounds, ear infections, urinary tract infections, intra-abdominal infections and respiratory tract infections, as well as hospital environments. A smaller fraction of the isolates, 5.6%, were isolated from environmental samples, including water and industrial sources (Fig. 1c). The presence of *P. aeruginosa* in such a variety of environmental niches illustrates the reported versatility of this organism. Although the distribution of sequences is heavily skewed towards clinical isolates, overall, the data suggest that *carP* homologues are present in strains residing in both clinical and environmental settings, suggesting its potential importance for *P. aeruginosa* survival in diverse ecological niches.

Second, we aligned the nucleotide sequences of all 2269 homologues in group 1 and for each nucleotide, we determined the number of homologues with altered sequence. In Fig. 3a, the bars indicate the percent of sequences (out of 2269) that have a nucleotide change in comparison to the aligned nucleotide in the *carP* sequence. The majority of the identified sequence variations occurred in the proximity of 5′ and 3′ ends of the gene and within the phytase/SdiA-regulated domain. Overall, for the entire length of *carP* sequence, we detected 155 nucleotide mutations, 42 of which are missense mutations (Fig. 3b). When selection was estimated on a per-codon basis, 24 of the missense mutations showed normalized d*N*-d*S* values higher than 1, indicating a positive selection (Fig. 3b, Table S1, available in the online version of this article). Using the alignment of the 2269 homologues in group 1, we also compared the rates of non-synonymous versus synonymous mutations over the entire full-length *carP* gene, as opposed to the regions representing the transmembrane helix (TMH) (bases 68–129), and the SdiA regulated (bases 187–942) domains to identify positive selection regions. Only the SdiA-regulated region seemed to be under a positive selection (*P*-value=0.01) indicating an advantageous mutation.

Then, we grouped the homologues with sequence variations according to their sources: clinical or environmental (Fig. 4). A total of 39 amino acid mutations were detected in CarP homologues harboured by 1367 clinical isolates and 13 mutations were detected in 60 environmental isolates. Considering the total number of sequences, mutations in the environmental strains occur five times more frequently (13/60=0.2) than in the clinical isolates (54/1367=0.04). The overall distribution of the mutations within the sequence, however, were similar in the two groups. The functional significance of these mutations will be the focus of further studies.

Third, we clustered 2248 selected full-length sequences, and the representative sequences of the 26 obtained clusters were aligned to construct a maximum-likelihood phylogenetic tree with PA0319 as an outgroup (Fig. 5). Considering the phylogenetic relationship and the isolation sources, the clusters fell into three groups, which showed distinct profiles of mutations. Interestingly, 40% of sequences in the first group originated from CF isolates, compared to about 8 and 5% of those in the second, and third groups, respectively. This pattern was reversed for hospital isolates that represented 13, 20 and 38% of the sequences in the first, second and third group, respectively. The mutation profiles showed a smaller number of mutations identified in the first group of sequences that increased substantially in the second, and particularly, the third group. It is worth noting, that none of the mutations identified in this first group showed a normalized d*N*-d*S* values higher than 1 in our per-codon selection analysis.

**Table 4.** Alterations in 2D structure of CarP and its domains due to identified mutations

| Protein/domain (no. of mutations) | α-Helix | B-sheets | Loops |
|---|---|---|---|
| Wild type CarP | 249 (77.6%) | 129 (40.2%) | 42 (13.1%) |
| Mutations within TMH domain (6) | 254 (79.1%) | 127 (39.6%) | 44 (13.7%) |
| Mutations within phytase-like domain (23) | 264 (82.2%) | 119 (37.1%) | 43 (13.4%) |
| All CarP mutations (40) | 267 (83.2%) | 165 (51.4%) | 44 (13.7%) |

**Table 5.** Sequence identity of *carP* homologues amplified from pooled CF isolates, each pool representing one patient*

| CF Sputum Sample ID | Regions | | | | Total coverage |
|---|---|---|---|---|---|
| | Age | TMH | SdiA regulated | Phytase-like | |
| 5814A | 6 | 100% | 99.74% | 99.83% | 99.79% |
| 102314A | 7 | 100% | 99.07 | 99.17% | 99.28% |
| 2515D | 17 | 100% | 100% | 99.83% | 100% |
| 6514D | 20 | 100% | 99.74% | 99.83% | 99.79 |
| 61214C | 22 | 96.77% | 99.74% | 99.83% | 99.48% |
| 31314A | 28 | 100% | 100% | 99.83% | 100.00% |
| 22014D | 29 | 100% | 99.60% | 99.83% | 99.28% |
| 22714D | 31 | 100% | 99.74% | 99.67% | 99.79% |
| 22014B | 39 | 98.39% | 99.74% | 99.83% | 99.69% |
| 6514C | 55 | 100% | 99.03% | 98.42% | 99.35% |

*Percentages represent the nucleotide sequence identity within the corresponding region of the gene in comparison to PAO1 *carP*.

Considering that CF isolates represent the largest number of sequences (40%) in the largest first group (*n*=1559), these data suggest the highest conservation of *carP* among CF isolates. This may indicate a particular importance of *carP* during *P. aeruginosa* infections in the lungs of these patients. The results, however, may be skewed by unequal representation of *P. aeruginosa* isolates from other sources.

## Conservation of CarP functional domains

To analyse whether the identified mutations have a potential to affect CarP functionality, we have mapped the nucleotide changes onto the predicted domains: a transmembrane helix (TMH) (bases 68–129) predicted to anchor CarP to the cytoplasmic membrane, SdiA regulated (bases 187–942) and phytase-like domain (bases 211–816) (Fig. 3). We also determined the percent identity within the domains in comparison to the wild-type *carP* in homologous sequences grouped according to their isolation source (Table 3). Overall, the highest sequence variation was observed within the TMH domain, where the sequence identity varied from 90–100% (Table 3). The overlapping phytase and SdiA-regulated domains showed the lowest sequence variation with a minimum sequence identity of 98.7%. Interestingly, the lowest sequence identity was observed for TMH domain

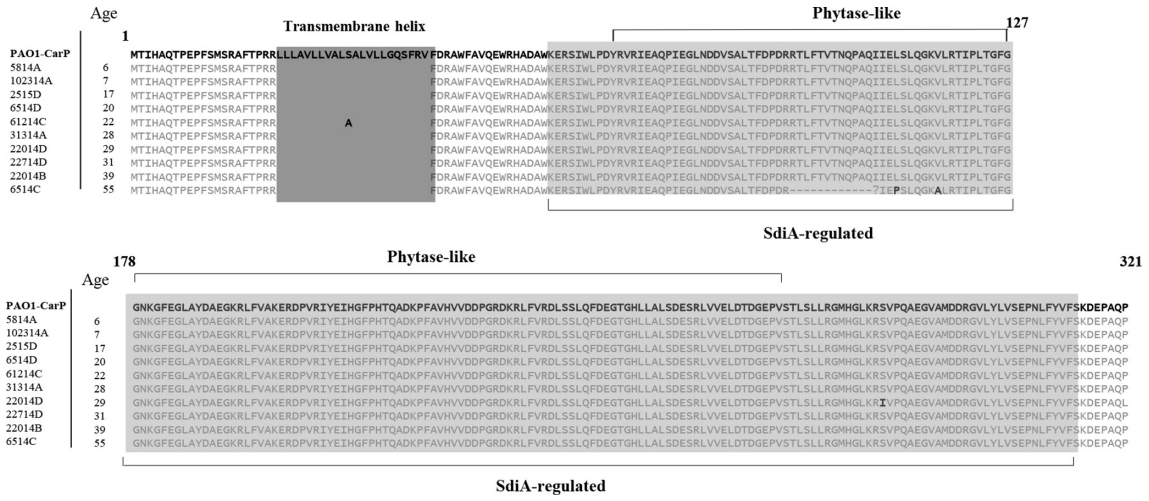



**Fig. 7.** Conservation of *carP* in *P. aeruginosa* isolates from sputa samples of CF patients. The *carP* conservation during CF infection was examined by sequencing ten pooled cultures, each representing a CF patient and including 96 randomly selected isolates. Sequence alignment was performed using MEGA [39]. 1. PAO1 *carP*, 2. 5814A (age 6), 3. 102314A (age 7), 4. 2515D (age 17), 5. 6514D (age 20), 6. 61214C (age 22), 7. 31314A (age 28), 8. 22014D (age 29), 9. 22714D (age 31), 10. 22014B (age 39) and 11 6514C (age 55). Only the regions of the gene that have mutations are shown. The TMH is shaded in dark grey and the protein domains are shaded in light grey.

(90%) and phytase/SdiA-regulated domains (98.7%) in two isolates from ear infections. More sequencing for isolates from diverse infections is required for identifying significant patterns in mutation rates.

To test whether the identified mutations would collectively affect the integrity of TMH domain, we introduced all six missense mutations (L23P, V26M, S32A, V35A, G38S and S40N) detected within this domain into the CarP protein sequence *in silico* and tested the prediction of the TMH domain. While the probability score was lower in the mutated protein (0.89 vs 0.98), the TMH domain was still predicted. We also compared the overall hydrophobicity, the transmembrane tendency, and the secondary structure of the *in silico* mutated TMH domain to that of CarP. Again, the differences were not substantial (Fig. S1), suggesting that these amino acid changes in CarP homologues do not impact the localization of the protein.

### Identified mutations do not impact the *β*-propeller fold of the protein

A large number of proteins with diverse functions form *β*-propeller structures that are well conserved among all branches of life (reviewed in [53]). Mutations disrupting the *β*-propeller structure commonly inactivate these proteins [54]. To test whether the 40 amino acid mutations identified within the phytase-like domain among 2269 CarP homologues alter the secondary structure and *β*-propeller fold of the protein, we introduced these mutations into CarP *in silico* and predicted the secondary and 3D structures of the mutated protein. The secondary structure prediction showed only a 5% gain in α-helixes and a 3% loss in *β*-sheets, and no structural differences were observed within the predicted loop regions (Table 4). The alignment of the predicted 3D structures of the mutated and wild-type CarP (Fig. 6) showed that only two mutations (P93S and K232R) located within blade I and blade V, respectively, altered the *β*-propeller fold, but no major changes in the predicted 3D structure were observed. Future studies will investigate the impact of these mutations on the function of CarP.

### *carP* was only detected in *P. aeruginosa* strains

Since the presented *in silico* analyses concluded that *carP* is primarily present and highly conserved in *P. aeruginosa,* we hypothesized that *carP* has potential as a biomarker. To test this, we designed primers specific to *carP*. Within 2269 *carP* homologues identified among *P. aeruginosa* strains, the 5′ region complimentary to the forward primer showed 0.04–0.97% mutation rate. The 3′ region recognized by the reverse primer showed even a lower occurrence of mutation, ranging from 0.66–0.79% sequences (Fig. S2).

PCR analyses using the designed *carP* primers showed a consistent amplification of *carP* in PAO1, but did not amplify any product from seven different bacterial genera, including *Escherichia*, *Staphylococcus, Micrococcus, Bacillus, Alcaligenes, Arthrobacteria* and *Klebsiella* (Table 1). *16S rRNA* primers were used as a positive control for each strain. Further, no amplification of *carP* was observed in 11 environmental *Pseudomonas* species, excluding *P. aeruginosa* (Fig. S3). These included nine strains (shared by Dr Mavrodi) isolated from plants, soil, streams or the rhizosphere and represented eight species: *P. protegens, P. fluorescens, P. chlororaphis, P. veronii, P. aridius, P. mosselii, P. synsaxatha* and *P. brassicacearum*. In addition, no *carP* amplification was observed for nine environmental isolates collected from soil, ponds and a laboratory sink (Fig. S3). They included *P. monteilii* (Table 1) and eight isolates morphologically drastically distinct from *Pseudomonas* (data not shown).

To show that the *carP*-specific primes can be used to detect *P. aeruginosa* strains in clinical samples, we tested a variety of laboratory and clinical isolates. Positive amplification of *carP* was obtained for all 60 tested *P. aeruginosa* strains, including laboratory, clinical and environmental isolates (Table 1, Fig. S3). Despite the potential impact of the mutations detected within the 5′ and 3′ regions of the gene on recognition by the primers, all the *P. aeruginosa* isolates tested produced amplicons, supporting our confidence in the designed primers.

Finally, we aimed to explore the conservation of *carP* sequences among CF sputa isolates collected at the clinic from ten patients of different ages ranging from 6 to 55. For this, for each of the ten patients, we pooled either 96 individual *Pseudomonas* isolates or the entire PIA plate of colonies, extracted their DNA, amplified by using *carP* specific primers, and sequenced the amplicons (Table 1, Fig S3). Considering the amplicons from mixed cultures were sequenced, to identify all possible variations in the sequences, the spectra were analysed closely for multiple calls of nucleotides potentially representing sequences of different strains. Similar to the sequences retrieved from the *Pseudomonas* database, most sequence variations occurred within the TMH domain with sequence identity dropping to 96.8 and 98.4% in two of the samples (Table 5). The rest of the sequence showed similar conservation to that of previously sequenced CF isolates. Only four of the mutations altered the protein sequence (Fig. 7). No correlation was observed between the number of mutations and the age of the patients. A larger sample pool is required for such correlative analysis.

Overall, these data illustrate the conservation of *carP* sequence in *P. aeruginosa* that presents an opportunity for this gene to serve as a biomarker for detecting *P. aeruginosa* in clinical and environmental samples.

Funding information
The research was supported by the NIH COBRE P20GM103648.

Acknowledgements
We thank Dr Michelle Callegan, University of Oklahoma Health Sciences for sharing the *Pseudomonas* ocular infection samples. We thank Drs Olga and Dmitri Mavrodi for sharing environmental isolates of *P. aeruginosa*. We thank Jodie Fox and Dr Nighat Mehdi at OU Children's Hospital in Oklahoma City for the collection of CF sputa samples and Elizabeth Pascual for the transport and recovery of the *P. aeruginosa* isolates. The Oklahoma Department of Health provided four clinical isolates. We would like to thank Rosalie Dohmen for her assistance with the extraction of DNA and PCR. We thank Dr Rolf Prade for critical

analysis and Mackenzie Hull for testing the *carP* primer amplification from non-Pseudomonads. Sequencing was performed at the OSU DNA Sequencing in Core Facility. The research was supported by the NIH COBRE P20GM103648.

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

### Ethical statement
Sputa samples were isolated during routine monitoring of the CF patients at the OU Children's Hospital, Oklahoma City, OK. Institutional Review Board at the University of Oklahoma approved sample collection by clinical staff upon patient consent. Allowed information about patients included age, gender, and whether a patient was experiencing an exacerbation at the time of sample collection.

### References

1. Bhagirath AY, Li Y, Somayajula D, Dadashi M, Badr S *et al*. Cystic fibrosis lung environment and *Pseudomonas aeruginosa* infection. *BMC Pulm Med* 2016;16:174.
2. Busi Rizzi E, Schininà V, Bordi E, Buontempo G, Narciso P *et al*. HIV-related bronchopulmonary infection by *Pseudomonas aeruginosa* in the HAART era: radiological findings. *Acta Radiol* 2006;47:793–797.
3. Stapleton F, Carnt N. Contact lens-related microbial keratitis: how have epidemiology and genetics helped us with pathogenesis and prophylaxis. *Eye* 2012;26:185–193.
4. Lamas Ferreiro JL, Álvarez Otero J, González González L, Novoa Lamazares L, Arca Blanco A *et al*. *Pseudomonas aeruginosa* urinary tract infections in hospitalized patients: Mortality and prognostic factors. *PLoS One* 2017;12:e0178178.
5. CDC. *Pseudomonas aerugionsa*, 2020; 2017.
6. Meyer JM, Neely A, Stintzi A, Georges C, Holder IA. Pyoverdin is essential for virulence of *Pseudomonas aeruginosa*. *Infect Immun* 1996;64:518–523.
7. Laarman AJ, Bardoel BW, Ruyken M, Fernie J, Milder FJ *et al*. *Pseudomonas aeruginosa* alkaline protease blocks complement activation via the classical and lectin pathways. *J Immunol* 2012;188:386–393.
8. Kang D, Kirienko DR, Webster P, Fisher AL, Kirienko NV. Pyoverdine, a siderophore from *Pseudomonas aeruginosa*, translocates into *C. elegans*, removes iron, and activates a distinct host response. *Virulence* 2018;9:804–817.
9. Look DC, Stoll LL, Romig SA, Humlicek A, Britigan BE *et al*. Pyocyanin and its precursor phenazine-1-carboxylic acid increase IL-8 and intercellular adhesion molecule-1 expression in human airway epithelial cells by oxidant-dependent mechanisms. *J Immunol* 2005;175:4017–4023.
10. Mulcahy LR, Isabella VM, Lewis K. *Pseudomonas aeruginosa* biofilms in disease. *Microb Ecol* 2014;68:1–12.
11. Percival SL, Suleman L, Vuotto C, Donelli G. Healthcare-associated infections, medical devices and biofilms: risk, tolerance and control. *J Med Microbiol* 2015;64:323–334.
12. Clapham DE. Calcium signaling. *Cell* 2007;131:1047–1058.
13. Ratner D, Mueller C. Immune responses in cystic fibrosis: are they intrinsically defective? *Am J Respir Cell Mol Biol* 2012;46:715–722.
14. Sarkisova S, Patrauchan MA, Berglund D, Nivens DE, Franklin MJ. Calcium-Induced virulence factors associated with the extracellular matrix of mucoid *Pseudomonas aeruginosa* biofilms. *J Bacteriol* 2005;187:4327–4337.
15. Guragain M, King MM, Williamson KS, Pérez-Osorio AC, Akiyama T *et al*. The *Pseudomonas aeruginosa* PAO1 two-component regulator CarSR regulates calcium homeostasis and calcium-induced virulence factor production through its regulatory targets CarO and CarP. *J Bacteriol* 2016;198:951–963.
16. Khanam S, Guragain M, Lenaburg DL, Kubat R, Patrauchan MA. Calcium induces tobramycin resistance in *Pseudomonas aeruginosa* by regulating RND efflux pumps. *Cell Calcium* 2017;61:32–43.
17. Guragain M, Lenaburg DL, Moore FS, Reutlinger I, Patrauchan MA. Calcium homeostasis in *Pseudomonas aeruginosa* requires multiple transporters and modulates swarming motility. *Cell Calcium* 2013;54:350–361.
18. Mena A, Smith EE, Burns JL, Speert DP, Moskowitz SM *et al*. Genetic adaptation of *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients is catalyzed by hypermutation. *J Bacteriol* 2008;190:7910–7917.
19. Wang Y, Gao L, Rao X, Wang J, Yu H *et al*. Characterization of lasR-deficient clinical isolates of *Pseudomonas aeruginosa*. *Sci Rep* 2018;8:13344.
20. Cornforth DM, Dees JL, Ibberson CB, Huse HK, Mathiesen IH *et al*. *Pseudomonas aeruginosa* transcriptome during human infection. *Proc Natl Acad Sci U S A* 2018;115:E5125–E5134.
21. Salma R, Dabboussi F, Kassaa I, Khudary R, Hamze M, Dabboussi F. gyrA and parC mutations in quinolone-resistant clinical isolates of *Pseudomonas aeruginosa* from Nini Hospital in north Lebanon. *J Infect Chemother* 2013;19:77–81.
22. De Vos D, Lim A, Pirnay JP, Struelens M, Vandenvelde C *et al*. Direct detection and identification of *Pseudomonas aeruginosa* in clinical samples such as skin biopsy specimens and expectorations by multiplex PCR based on two outer membrane lipoprotein genes, oprl and oprL. *J Clin Microbiol* 1997;35:1295–1299.
23. DA Silva Filho LVF, Levi JE, Bento CNO, DA Silva Ramos SRT, Rozov T. PCR identification of *Pseudomonas aeruginosa* and direct detection in clinical samples from cystic fibrosis patients. *J Med Microbiol* 1999;48:357–361.
24. Khan AA, Cerniglia CE. Detection of *Pseudomonas aeruginosa* from clinical and environmental samples by amplification of the exotoxin A gene using PCR. *Appl Environ Microbiol* 1994;60:3739–3745.
25. Žukovskaja O, Agafilushkina S, Sivakov V, Weber K, Cialla-May D *et al*. Rapid detection of the bacterial biomarker pyocyanin in artificial sputum using a SERS-active silicon nanowire matrix covered by bimetallic noble metal nanoparticles. *Talanta* 2019;202:171–177.
26. Kviatkovski I, Shushan S, Oron Y, Frumin I, Amir D *et al*. Smelling *Pseudomonas aeruginosa* infections using a whole-cell biosensor - An alternative for the gold-standard culturing assay. *J Biotechnol* 2018;267:45–49.
27. van Oort PM, Brinkman P, Slingers G, Koppen G, Maas A *et al*. Exhaled breath metabolomics reveals a pathogen-specific response in a rat pneumonia model for two human pathogenic bacteria: a proof-of-concept study. *Am J Physiol Lung Cell Mol Physiol* 2019;316:L751–L756.
28. Tian Y, Zeng T, Tan L, Wu Y, Yu J *et al*. Clinical significance of BPI-ANCA detecting in COPD patients with *Pseudomonas aeruginosa* colonization. *J Clin Lab Anal* 2019:e22908.
29. Emerson J, Rosenfeld M, McNamara S, Ramsey B, Gibson RL. *Pseudomonas aeruginosa* and other predictors of mortality and morbidity in young children with cystic fibrosis. *Pediatr Pulmonol* 2002;34:91–100.
30. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA *et al*. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* 2008;74:2461–2470.
31. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA *et al*. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 2009;6:343–.
32. Schneider CA, Rasband WS, Eliceiri KW. Nih image to ImageJ: 25 years of image analysis. *Nat Methods* 2012;9:671–675.
33. Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA *et al*. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res* 2016;44:D646–D653.
34. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
35. 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;22:4673–4680.

36. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database.. *J Bioinformatics* 2018.
37. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL *et al*. KBase: the United States department of energy systems biology Knowledgebase. *Nat Biotechnol* 2018;36:566–569.
38. Liu M, Li X, Xie Y, Bi D, Sun J *et al*. Iceberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res* 2019;47:D660–D665.
39. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870–1874.
40. King E. Isolate-Search. *GitHub* 2018.
41. Okonechnikov K, Golosova O, Fursov M, UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 2012;28:1166–1167.
42. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 2010;26:680–682.
43. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 1992;8:275–282.
44. Kumar TA. CFSSP: Chou and Fasman secondary structure prediction server. *J Wide Spectrum* 2013;1:15–19.
45. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* 2015;43:W174–W181.
46. DeLano WL. Pymol: an open-source molecular graphics tool. *CCP4 Newsletter on Protein Crystallography* 2002;40:82–92.
47. Pond SLK, Muse SV. HyPhy: hypothesis testing using phylogenies. *Statistical Methods in Molecular Evolution*. Springer; 2005. pp. 125–181.
48. Woodman M. Direct PCR of intact bacteria (colony PCR). *Curr Protoc Microbiol* 2008;9:A. 3D. 1-A. 3D. 6–.
49. Slack FJ, Ruvkun G. A novel repeat domain that is often associated with ring finger and B-box motifs. *Trends Biochem Sci* 1998;23:474–475.
50. Kawalek A, Kotecka K, Modrzejewska M, Gawor J, Jagura-Burdzy G *et al*. Genome sequence of *Pseudomonas aeruginosa* PAO1161, a PAO1 derivative with the ICEPae1161 integrative and conjugative element. *BMC Genomics* 2020;21:14.
51. Lotlikar SR, Hnatusko S, Dickenson NE, Choudhari SP, Picking WL *et al*. Three functional $\beta$-carbonic anhydrases in Pseudomonas aeruginosa PAO1: role in survival in ambient air. *Microbiology* 2013;159:1748–1759.
52. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF *et al*. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 2013;41:D991–D995.
53. Chen CK-M, Chan N-L, Wang AH-J. The many blades of the $\beta$-propeller proteins: conserved but versatile. *Trends Biochem Sci* 2011;36:553–561.
54. Laskowski RA, Tyagi N, Johnson D, Joss S, Kinning E *et al*. Integrating population variation and protein structural analysis to improve clinical interpretation of missense variation: application to the WD40 domain. *Hum Mol Genet* 2016;25:927–935.
55. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P *et al*. Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature* 2000;406:959–964.
56. He J, Baldini RL, Déziel E, Saucier M, Zhang Q *et al*. The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proc Natl Acad Sci U S A* 2004;101:2530–2535.
57. Silo-Suh LA, Suh S-J, Ohman DE, Wozniak DJ, Pridgeon JW. Complete genome sequence of *Pseudomonas aeruginosa* mucoid strain FRD1, Isolated from a cystic fibrosis patient. *Genome Announc* 2015;3:e00153–00115.
58. Paulsen IT, Press CM, Ravel J, Kobayashi DY, Myers GSA *et al*. Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat Biotechnol* 2005;23:873–878.
59. Loper JE, Hassan KA, Mavrodi DV, Davis EW, Lim CK *et al*. Comparative genomics of plant-associated Pseudomonas spp.: insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet* 2012;8:e1002784.
60. Biessy A, Novinscak A, Blom J, Léger G, Thomashow LS *et al*. Diversity of phytobeneficial traits revealed by whole-genome analysis of worldwide-isolated phenazine-producing *Pseudomonas* spp. *Environ Microbiol* 2019;21:437–455.
61. Wang X, Mavrodi DV, Ke L, Mavrodi OV, Yang M *et al*. Biocontrol and plant growth-promoting activity of rhizobacteria from Chinese fields with contaminated soils. *Microb Biotechnol* 2015;8:404–418.
62. Mavrodi DV, Ksenzenko VN, Bonsall RF, Cook RJ, Boronin AM *et al*. A seven-gene locus for synthesis of phenazine-1-carboxylic acid by *Pseudomonas fluorescens* 2-79. *J Bacteriol* 1998;180:2541–2548.
63. Lotlikar SR, Gallaway E, Grant T, Popis S, Whited M *et al*. Polymeric composites with silver (I) Cyanoximates inhibit biofilm formation of gram-positive and gram-negative bacteria. *Polymers* 2019;11:1018 [Epub ahead of print 09 Jun 2019].
64. Relman DA, Schmidt TM, MacDermott RP, Falkow S. Identification of the uncultured bacillus of Whipple's disease. *N Engl J Med* 1992;327:293–301.
65. Turner S, Pryer KM, Miao VP, Palmer JD. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J Eukaryot Microbiol* 1999;46:327–338.
66. Sievers F, Higgins DG. Clustal omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* 2014;1079:105–116.
67. Yang J, Yan R, Roy A, Xu D, Poisson J *et al*. The I-TASSER suite: protein structure and function prediction. *Nat Methods* 2015;12:7–.


Edited by: D. Grainger and M. Welch