Pseudomonas aeruginosa is one of the most common pathogens related to healthcare-associated infections. The Brazilian isolate, named CCBH4851, is a multidrug-resistant clone belonging to the sequence type 277. The antimicrobial resistance mechanisms of the CCBH4851 strain are associated with the presence of the blaSPM-1 gene, encoding a metallo-beta-lactamase, in combination with other exogenously acquired genes. Whole-genome sequencing studies focusing on emerging pathogens are essential to identify key features of their physiology that may lead to the identification of new targets for therapy. Using both Illumina and PacBio sequencing data, we obtained a single contig representing the CCBH4851 genome with annotated features that were consistent with data reported for the species. However, comparative analysis with other Pseudomonas aeruginosa strains revealed genomic differences regarding virulence factors and regulatory proteins. In addition, we performed phenotypic assays that revealed CCBH4851 is impaired in bacterial motilities and biofilm formation. On the other hand, CCBH4851 genome contained acquired genomic islands that carry transcriptional factors, virulence and antimicrobial resistance-related genes. Presence of single nucleotide polymorphisms in the core genome, mainly those located in resistance-associated genes, suggests that these mutations may also influence the multidrug-resistant behavior of CCBH4851. Overall, characterization of Pseudomonas aeruginosa CCBH4851 complete genome revealed the presence of features that strongly relates to the virulence and antibiotic resistance profile of this important infectious agent.
The state of São Paulo. Over the past years, several isolates carrying the bla gene identified in a clinical isolate recovered from a blood culture of a patient admitted to a hospital located in the state of São Paulo. Over the past years, several isolates carrying the bla gene were recovered from multiple P. aeruginosa hospital infection outbreaks widespread in the Brazilian territory. Multilocus sequence typing included most of these isolates within the sequence type 277 (ST-277) which contains MDR strains from different countries. However, bla presence was only detected in isolates originating from Brazil and it is unknown why SPM-1 is restricted to a specific region whereas other MBLs tend to spread worldwide. Nevertheless, the risk of SPM-1 worldwide dissemination should not be neglected. In 2008, an isolate was recovered from the catheter tip of a patient admitted in a hospital located in the state of Goiás. The strain, named CCBH4851, was subjected to antimicrobial susceptibility assays. Among the agents tested, the strain was resistant to aztreonam, amikacin, gentamicin, ceftazidime, cefepime, ciprofloxacin, imipenem, meropenem, and piperacillin-tazobactam, being susceptible only to polymyxin B. Multidrug resistance scenarios like those presented by CCBH4851 highlight the urgent need for the discovery of new therapeutic options. Thereby, recent whole-genome sequencing studies have focused on emerging MDR P. aeruginosa strains in order to identify key aspects of their physiology that may lead to the identification of new targets for drug development. This study was designed to thoroughly characterize the complete closed genome of P. aeruginosa CCBH4851 through the detection of genomic features, genome comparison with other P. aeruginosa strains, and phenotypic analysis. We believe this study will serve as a useful reference for future research using CCBH4851 as a model organism.

### Results

#### Genome features of *P. aeruginosa* CCBH4851 compared to *P. aeruginosa* PAO1 reference strain.

| Feature | CCBH4851 | PAO1 |
|---------|-----------|------|
| Genome status | Complete | Complete |
| Source | Catheter tip | Wound |
| Country | Brazil | Australia |
| Year | 2008 | Unknown |
| Genome size (bp) | 6,834,257 | 6,264,404 |
| G+C content (%) | 66.08 | 66.56 |
| No. of total genes | 6,342 | 5,697 |
| No. of pseudogenes | 86 | 19 |
| No. of total CDSs | 6,236 | 5,572 |
| No. of hypothetical proteins | 2,476 | 2,256 |
| No. of rRNAs | 13 | 13 |
| No. of tRNAs | 64 | 63 |
| No. of other RNAs | 29 | 30 |
| G+C content (%) | 66.08 | 66.56 |
| Genome size (bp) | 6,834,257 | 6,264,404 |
| Year | 2008 | Unknown |

MBLs have no activity only against monobactams but are not affected by SBL inhibitors. Although MBLs can be inactivated by chelating compounds, such as ethylenediaminetetraacetic acid, there is not known inhibitor available for clinical use. In Brazil, a recent report shows 42.9% of *P. aeruginosa* isolates recovered from primary bloodstream infections among adult patients hospitalized in intensive care units were resistant to carbapenems. Currently, the main carbapenem resistance mechanism associated with Brazilian isolates is the production of an MBL denominated São Paulo metallo-beta-lactamase (SPM-1). In 2002, the SPM-1-encoding gene (blaSPM-1) was first identified in a clinical isolate recovered from a blood culture of a patient admitted in a hospital located in the state of São Paulo. Over the past years, several isolates carrying blaSPM-1 were recovered from multiple *P. aeruginosa* hospital infection outbreaks widespread in the Brazilian territory. Multilocus sequence typing included most of these isolates within the sequence type 277 (ST-277) which contains *P. aeruginosa* strains from different countries. However, blaSPM-1 presence was only detected in isolates originating from Brazil and it is unknown why SPM-1 is restricted to a specific region whereas other MBLs tend to spread worldwide. Nevertheless, the risk of SPM-1 worldwide dissemination should not be neglected. In 2008, an isolate was recovered from the catheter tip of a patient admitted in a hospital located in the state of Goiás. The strain, named CCBH4851, was subjected to antimicrobial susceptibility assays. Among the agents tested, the strain was resistant to aztreonam, amikacin, gentamicin, ceftazidime, cefepime, ciprofloxacin, imipenem, meropenem, and piperacillin-tazobactam, being susceptible only to polymyxin B. Multidrug resistance scenarios like those presented by CCBH4851 highlight the urgent need for the discovery of new therapeutic options. Thereby, recent whole-genome sequencing studies have focused on emerging MDR *P. aeruginosa* strains in order to identify key aspects of their physiology that may lead to the identification of new targets for drug development. This study was designed to thoroughly characterize the complete closed genome of *P. aeruginosa* CCBH4851 through the detection of genomic features, genome comparison with other *P. aeruginosa* strains, and phenotypic analysis. We believe this study will serve as a useful reference for future research using CCBH4851 as a model organism.

#### Core, accessory and unique genomes

The gene repertoire of CCBH4851 was compared to other strains divided in three groups (Table 4): (A) composed by the commonly used reference strains in research studies, (B) composed by other ST-277 strains also carrying the blaSPM-1 gene, and (C) composed by other MDR strains belonging to different STs. A search for orthologs among these genomes identified genes that were shared by all genomes (core), genes specific to each genome (unique) or genes shared by two or more (but not all) genomes (accessory). In group A, 5,068 orthologous genes were identified as the core genome, 251 as the accessory genome and 697 genes were unique of CCBH4851 (Supplementary Table S1 online). Apart from genes included in the “Function unknown” COG class, the majority of unique genes were distributed among “Replication, recombination, and repair” and “Transcription” COG classes (Fig. 2A). In group B, the core genome comprised 5,645 genes. The accessory and unique genomes were comprised by 293 and 195 genes, respectively, also having the “Replication, recombination, and repair” and “Transcription” COG classes as the major functional categories with the exception of the “Function unknown” family (Fig. 2B). Group C, including MDR strains of different STs, presented 3,680 genes as the core genome, 1,745 as the accessory genome and 508 as unique genes of CCBH4851. The most evident functional categories remained the same (Fig. 2C). The comparative analysis

| Feature | CCBH4851 | PAO1 |
|---------|-----------|------|
| Year | 2008 | Unknown |
| Source | Catheter tip | Wound |
| Country | Brazil | Australia |
| Genomic size (bp) | 6,834,257 | 6,264,404 |
| G+C content (%) | 66.08 | 66.56 |
| No. of total genes | 6,342 | 5,697 |
| No. of pseudogenes | 86 | 19 |
| No. of total CDSs | 6,236 | 5,572 |
| No. of hypothetical proteins | 2,476 | 2,256 |
| No. of rRNAs | 13 | 13 |
| No. of tRNAs | 64 | 63 |
| No. of other RNAs | 29 | 30 |
also revealed that the CCBH4851 genome has 205 genes lacking or presenting partial homology when compared to the PAO1 genome (Supplementary Table S2 online). The predominant functional COG categories of the non-orthologous genes were “Replication, recombination, and repair”, “Transcription”, “Intracellular trafficking, secretion, and vesicular transport”, “Inorganic ion transport and metabolism”, “Energy production and conversion”, “Secondary metabolites biosynthesis, transport, and catabolism”, “Defense mechanisms”, and “Coenzyme transport and metabolism”. It is noteworthy that oprD (porin), mexZ, liuR, PA0207, PA0306, PA2100, PA2220, PA2547, PA3067, PA3094, and PA3508 (transcriptional regulators), opdE (membrane protein), pilA (type IV fimbrial protein), algP (alginate regulatory protein), and others were among these genes.

**Genomic islands and mobile elements.** The *P. aeruginosa* CCBH4851 genome contained 19 integrated genomic islands with a total of 705 CDSs (Fig. 1, yellow highlighted regions; Supplementary Table S3 online). Eighteen islands were homologous to *Pseudomonas aeruginosa* genome islands (PAGIs) previously found in other isolates belonging to ST-2777.* An additional region found by IslandViewer was less than 5,000 bp, but had a G+C content of 55.13% and encoded 5 proteins including an integrase and an helix-turn-helix (HTH) domain-containing protein. This region was annotated as a new PAGI, numbered PAGI-41. Overall, 69.92%
of the genes located in the PAGIs had no function assigned. The remainder were mainly classified in the following COG categories: "Replication, recombination, and repair", "Transcription", "Intracellular trafficking, secretion, and vesicular transport", "Energy production and conversion", "Secondary metabolism biosynthesis, transport, and catabolism", "Inorganic ion transport and metabolism" and "Defense mechanisms". A total of 13 insertion sequences (ISs) were predicted and listed in details in Supplementary Table S8 online.

The hypothetical structural effect of new synonymous variants, 4,972 missense variants, 7 stop gained, 5 stop lost, 5 start lost, 39 in frame deletions/insertions were common to a few functional categories or were randomly distributed among all of them. Mutations occurred mainly in genes belonging to "Inorganic ion transport and metabolism", "Amino acid transport and metabolism", "Secondary metabolites biosynthesis, transport, and catabolism" and "Cell wall, membrane, and envelope biogenesis", "Transport" and "Defense mechanisms".

Comparative genome analysis revealed strong evidences of horizontal gene transfer events in the CCBH4851 genome. Table 2 summarizes a comparison between CCBH4851 and PAO1 regulatory proteins. In accordance with the comparative analysis, the CCBH4851 genome lacks the histidine-kinase PA2583, the response regulator PA0034, as well as the transcriptional regulators PA3508, PA3067, and LiuR (PA2016). However, the CCBH4851 genome possessed 28 additional regulatory proteins, mostly transcriptional regulators with Xre- and LysR-type domains, distributed among PAGIs (Supplementary Table S5 online).

Single nucleotide polymorphisms (SNPs) detection between P. aeruginosa CCBH4851 and PAO1 genomes identified 25,220 variant types, including SNPs, multiple nucleotide polymorphisms (MNP), insertions, deletions and a combination of SNP/MNP. Results revealed 16,690 synonymous variants, 4,972 missense variants, 7 stop gained, 5 stop lost, 5 start lost, 39 in frame deletions/insertions, 40 frameshift variants, and 3440 variant types affecting intergenic regions (Supplementary Table S7 online). Synonymous variants were left aside, as they have no presumable impact on cellular processes. Apart from intergenic regions, the remaining affected genes were classified into COG families to assess whether variant types were common to a few functional categories or were randomly distributed among all of them. Mutations occurred mainly in genes belonging to "Inorganic ion transport and metabolism", "Amino acid transport and metabolism", "Cell wall, membrane, and envelope biogenesis", "Transcription", and "Energy production and conversion" COG categories (Fig. 3). In addition, Table 3 summarizes a list of virulence and antimicrobial resistance-associated genes with their respective amino acid substitutions. The hypothetical structural effect of new substitutions was predicted and listed in details in Supplementary Table S8 online.
Phylogenomic analysis of *P. aeruginosa* CCBH4851. The relatedness of CCBH4851 genome was assessed by phylogenetic inference amongst the isolates of group A, B, and C using the whole genome to perform a SNP calling and multiple sequence alignment of the resulting core genome. *P. aeruginosa* PAO1 was used as reference genome and had 92.9% of its sequence covered by all isolates. Interesting, isolates clustered according to the sequence types in the same phyletic clade (Fig. 4). Also, phylogenomic analysis showed that CCBH4851 was more nearly related to isolates of ST-244, sharing about 2-fold less SNP variants with them than with ST-235 isolates. *P. aeruginosa* PA7, as expect, was excluded from the resulting tree since it is known to be a taxonomic outlier10.

Swarming, swimming and twitching motilities. Swarming, swimming and twitching assays showed that CCBH4851 is impaired in all three types of motility. Swimming and swarming zones of CCBH4851 were about 5 millimeters (diameter) smaller when compared to PAO1 (Figs. 5 and 6, respectively). Both differences were statistically significant (swimming: *P* value = 0.0048, two-tailed, unpaired t test; swarming: *P* value < 0.001, two-tailed, unpaired t test). Comparisons of twitching motility showed no differences concerning the diameter of migration zones (Fig. 7A). However, after removal of unattached cells and staining with crystal violet, we observed diminished attachment of CCBH4851 cells to the polystyrene surface when compared to PAO1 (Fig. 7B). Observation of the stained area showed that CCBH4851 cells produced an expanding circular adherent zone, indicating that only cells on the outer side of the twitching area were attached, whereas PAO1 cells seemed to adhere to the polystyrene surface throughout the entire area, with only a few bacteria released from the center of the colony. Absorbance measurements of solubilized crystal violet confirmed the significant difference seen between the two strains. In order to assess if the diminished attachment observed would affect the ability of CCBH4851 to form biofilms, 96-well plate biofilm formation assays were performed. Under the conditions tested, CCBH4851 formed about fourfold less biofilm than PAO1 (Fig. 7C, *P* value < 0.0001, two-tailed, unpaired t test).

Antimicrobial susceptibility. Antimicrobial susceptibility testing was performed using one drug of each of four major antimicrobial classes. Tests were performed using the broth microdilution method, which allowed us to determine a minimal inhibitory concentration (MIC) for each drug. CCBH4851 was resistant to all antimicrobial agents tested except for colistin. Colistin susceptibility tests presented more than one skipped well, with no pattern. According to the European Committee on Antimicrobial Susceptibility Testing (EUCAST) guidelines, tests with more than one skipped wells should not be reported11. However, the majority of replicates showed no growth above the resistance breakpoint for colistin. The MICs for gentamicin, imipenem and ciprofloxacin were >1024 µg/mL, 128 µg/mL and 16 µg/mL, respectively. In addition, an E-test strip containing ceftolozane/tazobactam (a cephalosporin in combination with a beta-lactamase inhibitor) was also employed and CCBH4851 showed no inhibition zone (Supplementary Fig. S2 online).

**Discussion**

*Pseudomonas aeruginosa* CCBH4851 is a clinical isolate originating from Brazil, which belongs to the endemic group ST-277. Previous studies demonstrated that ST-277 strains have a highly conserved DNA sequence and share several virulence and antimicrobial resistance-related features. However, to date, the presence of the *blaOXA-1* gene conferring resistance to carbapenems is a singular trait of Brazilian clones6,7. The purpose of this study was to reassemble and to re-annotate the CCBH4851 genome in order to perform a thorough characterization and a genomic comparison with other *P. aeruginosa* strains. The reassembly resulted in a single contig representing the complete closed genome of CCBH4851. Together with the provided annotation, data such as chromosome size, G+C content, number of CDSs, structural RNAs, and tRNAs were consistent with general features reported for *P. aeruginosa* strains12. In addition, the CCBH4851 genome alignment with PAO1, PA14, and ST-277 strains revealed a strong synteny (Supplementary Fig. S1 online) indicating an accurate assembly and annotation.

The analyses performed in this work showed that the core genome of CCBH4851, PAO1, and PA14 strains (group A, Table 4) covered more than 80% of the CCBH4851 genome. A similar percentage was found when the core genome was defined comparing CCBH4851 with other isolates belonging to the ST-277 also carrying *blaOXA-1* (group B, Table 4). However, when comparing MDR strains belonging to different STs (group C, Table 4), the core genome was smaller and an increase in the number of genes comprised in the accessory genome was

| Strain  | Predicted regulatory proteins |
|---------|-----------------------------|
|         | TCS HK | TCS RR | TCS PP | TF TR | TF OCS | TF RR | TF SF | TF ODP |
| PAO1    | 62 73 4 | 170 185 | 49 24 36 |
| CCBH4851| 61 72 5 | 189 191 | 48 25 38 |

Table 2. Regulatory proteins of *P. aeruginosa* CCBH4851 compared to *P. aeruginosa* PAO1 reference strain. 

*a* A PP homologous sequence is present in the PAO1 DNA sequence, but it is not annotated in the deposited Genbank file. TCS two-component systems, TF transcription factors, ODP other DNA-binding proteins, HK histidine-kinases, RR response regulators, PP phosphotransferase proteins, TR transcriptional regulators, OCS one-component systems, SF sigma factors.
observed. This result could be attributed to different factors: (i) some strains used in this analysis had over one hundred contigs and this lack of continuity could affect genome annotation; (ii) the strains used in this analysis are from different countries and had distinct resistance profiles, which could suggest a variation of mutation patterns and acquired regions, depicting interstrain differences. The larger accessory genome among these strains suggests a variability in pathogenicity and environmental flexibility.

Table 3. SNPs found in virulence and antimicrobial resistance-associated genes of *P. aeruginosa* CCBH4851 using *P. aeruginosa* PAO1 as reference.

| Gene name | Amino acid substitution | Product |
|-----------|-------------------------|---------|
| *alg8*    | L61V                    | Glycosyltransferase Alg8 |
| *algC*    | T113A, R120L, A124V, L274R, T289A | Phosphomannomutase |
| *algE*    | N34G, T475A             | Alginase production protein AlgE |
| *algF*    | T11N                    | Alginase o-acetyltransferase AlgF |
| *algI*    | N403S, V1360AF          | Alginase o-acetylase AlgI |
| *algK*    | M202L                   | Alginase biosynthesis protein AlgK |
| *algX*    | Y49D                    | Alginase biosynthesis protein AlgX |
| *amgS*    | I260V                   | Protein AmgS |
| *ampC*    | R79Q*, T105A*           | Beta-lactamase |
| *ampD*    | G148A, S175L            | N-acetyl-anhydromuranmyl-L-alanine amidase |
| *ampP*    | M87L, T172A, M177V      | Transporter |
| *armZ*    | L88P, D161G, H182Q, V243A* | MexZ anti-repressor ArmZ |
| *dacB*    | A394P                   | D-allyl-D-alanine carboxypeptidase |
| *exoT*    | A83S, M107L, A135V, V216A | Exoenzyme T |
| *exoY*    | P203L, S366G            | Adenylate cyclase |
| *gltT*    | A439T                   | Glycerol-3-phosphate transporter |
| *gyrA*    | T83I*                   | DNA gyrase subunit A |
| *lasA*    | V192A, S223R            | Protease LasA |
| *lasB*    | S241G, S436L            | Elastase LasB |
| *lasR*    | V221G                   | Transcriptional regulator LasR |
| *mexD*    | E257Q, S845A            | Resistance-nodulation-cell division (RND) multidrug efflux transporter MexD |
| *mexE*    | P397Q, A407V            | Resistance-nodulation-cell division (RND) multidrug efflux membrane fusion protein MexE |
| *mexI*    | A782E                   | Resistance-nodulation-cell division (RND) efflux transporter |
| *mexS*    | D249N                   | Oxireductase |
| *mexT*    | F172I                   | Transcriptional regulator MexT |
| *mexZ*    | L174Q                   | Transcriptional regulator |
| *mucD*    | I137V                   | Serine protease MucD |
| *nalC*    | G71E*, S209R*           | Transcriptional regulator |
| *oprD*    | T103S, V1276*, F170L*   | Porin D |
| *oprE*    | DM688GV                 | Multidrug efflux outer membrane protein OprE |
| *oprM*    | A261T                   | Outer membrane protein OprM |
| *PA0032*  | A186V, L214V            | Transcriptional regulator |
| *PA1184*  | A249V                   | Transcriptional regulator |
| *PA2018*  | T543A*, Q840E*          | Multidrug efflux protein |
| *PA2019*  | K329Q*, L331V*, W358R*  | Multidrug Efflux lipoprotein |
| *PA3271*  | A212V                   | Two-component sensor |
| *parC*    | S87L*, H262Q            | DNA topoisomerase IV subunit A |
| *pmrA*    | L71R*                   | Two-component regulator system response regulator PmrA |
| *pmrB*    | Y345H*                  | Two-component regulator system signal sensor kinase PmrB |
| *rhlA*    | Q168P                   | Rhamnosyltransferase subunit A |
| *rhlC*    | E21K                    | Rhamnosyltransferase |
| *rhlI*    | D83E                    | Acyl-homoserine-lactone synthase |
| *rpoB*    | V51H                    | DNA-directed RNA polymerase subunit beta |
| *toxA*    | T4I, L14I, V55I, I432V* | Exotoxin A |
resistance mechanism, mainly when combined with overexpression of AmpC and efflux pump systems\(^1^3\). As a result, CCBH4851 showed a MIC of 128 \(\mu g/mL\) for imipenem.

Comparison with PAO1 also revealed the mutation of \(mexZ\) in CCBH4851 genome, suggesting the overexpression of at least one efflux pump system, MexXY, whose transcription is repressed by MexZ. A deletion of 17 nucleotides from position 439 in \(mexZ\) gene sequence caused a frameshift resulting in the loss of the stop codon.

**Figure 3.** Functionally clustered genes affected by SNPs causing amino acid substitutions. Description of COG categories: (C) energy production and conversion; (D) cell cycle control, cell division, chromosome partitioning; (E) amino acid transport and metabolism; (F) nucleotide transport and metabolism; (G) carbohydrate transport and metabolism; (H) coenzyme transport and metabolism; (I) lipid transport and metabolism; (J) translation, ribosomal structure, and biogenesis; (K) transcription; (L) replication, recombination, and repair; (M) cell wall, membrane, and envelope biogenesis; (N) cell motility; (O) post-translational modification, protein turnover, and chaperones; (P) inorganic ion transport and metabolism; (Q) secondary metabolites biosynthesis, transport, and catabolism; (S) function unknown; (T) signal transduction mechanisms; (U) intracellular trafficking, secretion, and vesicular transport; (V) defense mechanisms; (*) genes with two or more COG categories assigned; following the respective colors in the legend.

**Figure 4.** Relatedness of \(P.\ aeruginosa\) genomes. Core genome phylogenetic tree based on SNP calling using \(P.\ aeruginosa\) PAO1 as reference. Three distinct phyletic clades are highlighted corresponding to multilocus sequence typing: ST-277 (yellow shaded), ST-244 (red shaded), and ST-235 (blue shaded).
Mutations in mexZ are often present in clinical isolates overexpressing MexXY. Aminoglycoside resistance is known to be produced not only by the overexpression of MexXY but the presence of aminoglycoside-modifying enzymes (AMEs) and rRNA methylases. In addition to the possible efflux mechanism, CCBH4851 acquired...
**Figure 7.** The twitching phenotype of *P. aeruginosa* PAO1 and CCBH4851 strains. (a) Measurement of the circular zone diameter of PAO1 (green circles) and CCBH4851 (blue squares), and visual aspect of the respective twitching zone in 1% LB agar plates. (b) Measurements of solubilized crystal violet absorbance of PAO1 (green circles) and CCBH4851 (blue squares), and visual aspect of respective stained cells area. (c) Quantification of biofilm formation of PAO1 (green bar) and CCBH4851 (blue bar). Horizontal lines are the mean of all replicates. Error bars represent the standard deviation of the mean.

| Table 4. Bacterial strains used for comparative genome analysis. |
| --- |
| Strain | No. of contigs | Total size (bp) | Sequence type | Accession no. |
| **- Group A** | | | | |
| PAO1 | 1 | 6,264,404 | 549 | NC_002516 |
| UCBPP-PA14 (PA14) | 1 | 6,537,648 | 253 | NC_008463 |
| **- Group B** | | | | |
| PA1088 | 1 | 6,721,480 | 277 | CP015001 |
| PA3448 | 2 | 6,794,242 | 277 | LVWC01000000 |
| PA7790 | 1 | 7,018,690 | 277 | CP014999 |
| PA8281 | 1 | 6,928,736 | 277 | CP015002 |
| PA11803 | 1 | 7,008,578 | 277 | CP015003 |
| PA12117 | 5 | 6,645,782 | 277 | LVXB01000000 |
| **- Group C** | | | | |
| PA7 | 1 | 6,588,339 | 1195 | NC_009656 |
| NCGM2_S1 | 1 | 6,764,661 | 235 | AP012280 |
| WH-SGI-V-07170 | 119 | 6,813,449 | 235 | NZ_LLLZ00000000 |
| WH-SGI-V-07694 | 106 | 6,776,438 | 235 | NZ_LLSD00000000 |
| AZPZAE15021 | 129 | 6,686,424 | 244 | NZ_JTNZ00000000 |
| WH-SGI-V-07409 | 144 | 6,932,716 | 244 | NZ_LLOW00000000 |
| WH-SGI-V-07705 | 130 | 6,802,755 | 244 | NZ_LLSZ00000000 |
two AME-encoding genes, *aac(6’)-I* and *aadA7*, and a 16S rRNA methylase-encoding gene, *rmtD*, resulting in a MIC over 1024 µg/mL for gentamicin.

Other differences revealed by the comparative analysis were the mutations of *algP* and *pilA* genes, both related to biofilm development. AlgP is a histone-like protein which activates the transcription of *algD*, responsible for the alginate precursor synthesis. Alginate is an important molecule for the structural stability of the biofilm matrix. The amino acid sequence of AlgP contains several repeated KPAA motifs which are mutation targets at high frequency in clinical isolates. According to this, the *algP* gene sequence of CCBH4851 had an insertion of 12 nucleotides, adding one KPA motif to the AlgP protein sequence. The repeated 12-bp sequences of *algP* are considered a hot spot for DNA rearrangements and could provide a reversible switching mechanism between nonmucoid and mucoid phenotype, thus turning on and off virulence factors important for a successful infection16, 17. On the other hand, the *pilA* sequence is completely absent from the CCBH4851 genome. PilA is a pilin protein involved in type IV pilus (TFP) biogenesis, a major surface adhesin. Previous work demonstrated that the absence of *pilA* in carbenapenem-resistant clinical isolates is not uncommon18.

*P. aeruginosa* is capable of switching between three types of motility depending on the environment: swimming, swarming and twitching. Swimming is mediated by flagella in aqueous environments. Swarming is mediated by both flagella and TFP in semisolids environments. Twitching is TFP-dependent and allows the cells to move across surfaces by extension, attachment and retraction of the pilus filament. Also, surface attachment mediated by TFP is required for biofilm initiation19–21. As expected, our *in vitro* analyses showed that CCBH4851 is impaired in swimming, twitching, and biofilm development. *P. aeruginosa* CCBH4851 was also defective in its ability to swim. Although CCBH4851 presented no lack of known genes directly involved in flagellar biosynthesis, several SNPs were located in those genes (Supplementary Table S7 online), including *flhF, flhA, flaN, motB*, and others. During acute infection, motility is required to the colonization of environments, attachment to surfaces and initiation of biofilm formation, mechanisms strongly related to virulence. Motile bacteria often display a more virulent phenotype. Loss of motility is important to the switching for a sessile lifestyle and the establishment of chronic infection. Also, it is noteworthy that virulence and multidrug resistance are often inversely correlated, i.e. MDR strains are less or non-virulent whereas non-MDR strains are hypervirulent22, 23. In that sense, the phenotype displayed by CCBH4851 suggest a less virulent behavior in favor of a survival strategy under selective environmental pressures.

The acquisition of exogenous material by horizontal gene transfer is a common adaptive mechanism among *P. aeruginosa* strains, often related to the presence of genomic islands. Genomic islands are clusters of genes often encoding virulence factors, antimicrobial resistance proteins, toxins, secretion system proteins, transcriptional regulators, and other proteins24. The CCBH4851 genome presented a high number of islands, all sharing homology with previously described PAGIs, except for PAGI-41. Although the majority of proteins encoded by these PAGIs are hypothetical proteins, some of them could be clustered into functional categories due to the presence of conserved domains. The outstanding categories comprised genes involved in mechanisms of replication, recombination and repair, and transcription. Indeed, the newly annotated PAGI-41 carries the gene AL347_09405 which encodes an XRE-type HTH domain-containing protein, suggesting a function as a transcriptional repressor. Downstream, the gene AL347_09410 encodes a protein containing a RelE/ParE toxin superfamily domain. In response to environmental stress, such as antimicrobial exposure, this type of toxin can inhibit replication or translation mechanisms, leading the cell to a dormant state. Cells in dormant state, also called persister cells, become multidrug-tolerant. The repressor binds to the toxin forming an inactive complex acting as an antitoxin, and is able to bind its own promoter region preventing toxin expression. Antitoxins could also regulate transcription of several virulence factors, including biofilm formation25, 26. CCBH4851 possess three additional toxin-antitoxin system in other PAGIs. Moreover, all acquired genes conferring resistance to a broad range of antimicrobial agents were located in PAGIs, including the two copies of the carbapenem resistance gene *blaKPC*. Our *in vitro* analyses showed that CCBH4851 is not only resistant to several antimicrobial agents, but also presented MIC values over the resistance breakpoints.

In addition to the presence of PAGIs, distinct IS families were also detected in the CCBH4851 genome. The importance of ISs is not restricted to their role in horizontal gene transfer, but ISs movement along the chromosome can also affect antibiotic resistance by the activation of gene expression27. Another player in shaping the bacterial genome is the CRISPR-Cas system. The type I-C found in the CCBH4851 genome was previously described in other ST-277 bacteria as well as in members of ST-2357, 28. A recent study suggests the protective effect of CRISPR-Cas systems is more evident at the population level than at an evolutionary scale29. This could explain the intracranial genome conservation already observed among Brazilian isolates belonging to ST-277 in previous work3 and corroborated here by our phylogenomic analysis.

During the infection course, *P. aeruginosa* strains tend to adapt to the selective pressures of the environment, often through mutations in intergenic and/or coding sequence regions. A classical example of adaptive mutation is the MexT protein, responsible for activation of the *mexEF-OprN* operon. MexEF-OprN is quiescent in wild-type cells and its expression occurs in MexT or MexT-related genes. Indeed, the mexT gene sequence of CCBH4851 had an 8-bp deletion known to render MexT active, causing the induction of *mexEF-oprN* transcription and a decrease in OprD levels, which characterizes the so-called *nfxC*-type mutants30. However, later work showed that this 8-bp deletion alone was not sufficient to activate the transcription of *mexE*, and the generation of *nfxC*-type mutants seems to be multifactorial. Indeed, inactivation of *mexS*, a gene upstream of *mexT*, seems to be one of these additional factors31. The *mexS* gene of CCBH4851 suffered a deletion of 1 nucleotide in position 22 causing a frameshift which resulted in the loss of the stop codon. The gene product of the *mexS* pseudogene showed an alignment of only 7 aa with the MexS wild-type sequence of PAO1. This could suggest the absence of MexS in the CCBH4851 proteome, leading to the overexpression of *mexE* and other phenotypes related to *nfxC*-type mutants. Table 3 summarizes other SNPs in virulence and resistance-associated genes that are frequently mutated in *P. aeruginosa* clinical isolates. Among them, amino acid substitutions T831
in GyrA and S87L in ParC are well-known to play an important role in fluoroquinolone resistance as well as the overexpression of efflux pump systems\(^2\). Once again, the antimicrobial susceptibility testing revealed a MIC of 16 \(\mu g/mL\) for ciprofloxacin, corroborating the observed genotypes.

Quantification of efflux pump systems' expression is required to confirm altered transcription levels, but amino acid substitutions in NaLC (repressor of \(mexA\) gene), ArmZ (MexZ anti-repressor), PA2018 and PA2019 proteins suggest additional mechanisms leading to the overexpression of \(mexA\) and \(mexX\). In addition to these factors, amino acid substitutions R79Q and T105A found in AmpC (Table 3) characterizes this protein as a variant type named PDC-5. Clinical isolates carrying this variant presented AmpC overexpression, increased beta-lactamase activity and reduced susceptibility to ceftazidime, cefepime, ceftipirome, aztreonam, imipem and meropenem\(^{14,32-34}\). Although SPM-1 is an enzyme capable of hydrolyzing a broad range of beta-lactams, it is known that MBLs has little or no effect against monobactams\(^2\). However, CCBH4851 was resistant to aztreonam\(^4\), a monobactam widely available for clinical use. This phenotype can be explained by the presence of PDC-5, depicting a synergistic effect between AmpC and SPM-1 on beta-lactam resistance mechanisms of CCBH4851.

Not all SNPs found in this work were previously described; however, different mutations in the same genes listed in Table 3 were observed in other clinical isolates, suggesting a recurrence in the mutations' location caused by selective pressure. The variants found in CCBH4851 could be intraclonal, but could still change gene function, thus contributing to the enhanced resistance of ST-277 clones. In fact, the \(in\ \text{silico}\) analysis suggests some of these SNPs could affect protein function due to the introduction of amino acids with different properties, such as size, charge, hydrophobicity, often located in protein domains. The prediction indicates noteworthy mutations occurring in genes such as \(ampD\) (negative regulator of \(ampC\) expression), \(mucD\) (repressor of \(algU\) alternative sigma factor transcription), \(opf\) (member of \(mexCD\) efflux system), and others, which could affect protein folding and/or function (Supplementary Table S8 online).

The features found in the \(P.\ \text{aeruginosa}\) CCBH4851 complete genome are somehow related to its pathoadaptive behavior, since part of them are common among clinical isolates and some are unique of ST-277 clones. The number of unique genes, the number of genes contained in the acquired genomic islands, and the larger size of the CCBH4851 chromosome are consistent with this observation. However, most of the genome is shared with more susceptible strains, suggesting the high number of mutations found in conserved genes are contributing to the success of this MDR strain. Further validation of uncharacterized polymorphisms revealed by this study should help increase the understanding of MDR strains phenotype. The results provided in the present study will stimulate and contribute to future works that lead to the discovery of new antimicrobial targets and to the development of effective therapeutic choices against \(P.\ \text{aeruginosa}\) infections.

**Methods**

**Bacterial strains.** The focus of the present study was the bacterial strain \(Pseudomonas\ \text{aeruginosa}\) CCBH4851. The isolate is available in the Culture Collection of Hospital-Acquired Bacteria (CCBH), located at Fundação Oswaldo Cruz (WDCM947; CGEN022/2010). Other strains used to perform comparative genome analysis are listed in Table 4.

**Genome sequencing, reassembly and re-annotation.** Whole-genome sequencing of CCBH4851 was previously performed using the Illumina MiSeq platform\(^9\). Genome assembly resulted in a draft genome comprising 150 contigs, which were annotated as described in previous work. Later, an additional whole-genome sequencing was performed using the PacBio platform. Here, \(de\ \text{novo}\) assembly was carried out using the MaSuRCA assembler, which allowed the combination of Illumina short reads with PacBio long reads resulting in an improvement of the original assembly\(^{10}\). A single contig was obtained and was annotated using a customized pipeline. Briefly, annotation from the closely related strain \(P.\ \text{aeruginosa}\) PAO1 was transferred to the CCBH4851 genome using the Rapid Annotation Transfer Tool software\(^{35}\). Then, CDSs were predicted by GeneMark\(^\text{S}\)\(^\text{S}\)\(^\text{S}\). Prediction of rRNAs and tRNAs were performed by Rfam and tRNAscan SE software, respectively\(^{36,39}\). The predicted CDSs were functionally annotated based on homology searches against public databases, such as RefSeq\(^{40}\), COG\(^{41}\), KEGG\(^{42}\), UniProt\(^{43}\) and InterPro\(^{44}\). Finally, the annotations from previous steps were compared, combined and manually curated. Additionally, PRIAM software was used to assign an enzyme commission number to CDSs\(^{45}\). All genome visualization was made using the Artemis software\(^{46}\). The circular plot of CCBH4851 genome was produced by Circos software\(^{47}\).

**Comparative genome analysis.** Whole-genome comparison was performed to identify similarities and differences between CCBH4851 and each group of strains listed in Table 4. The genomes were analyzed using the bidirectional best-hit (BBH) clustering method based on homology searches using the BLAST algorithm to identify pairs of corresponding genes that are each other's best hit when different genomes are compared\(^{48}\). All against all BLAST alignments were performed using the following parameters: \(\geq 90\%\) coverage, \(\geq 90\%\) similarity and \(E\) value cut-off of 1e-10. A Python algorithm was applied over BLAST results to identify the BBHs\(^{49}\). Core, accessory, and unique genomes were analyzed using a MySQL database created with the data generated in previous steps. The CDSs were classified into Clusters of Orthologous Groups (COG) functional categories using the eggNOG-Mapper web application\(^{40}\). The identification of genomic islands was carried out by IslandViewer\(^{50}\). Insertion sequences were identified using the ISSaga web application\(^{51}\). Presence of a CRISPR-Cas system was assessed by the CRISPRCasFinder web application\(^{52}\). In order to detect genes involved with antimicrobial resistance mechanisms, protein sequences of the CCBH4851 genome were used to perform BLAST searches against CARD using the Resistance Gene Identifier web application\(^{53}\). Also, protein sequences were searched against the Virulence Factors Database of Pathogenic Bacteria\(^{54}\) using BLASTP with \(\geq 75\%\) coverage, \(\geq 50\%\) identity and \(E\) value \(\leq 1e^{-5}\).
value cut-off of 1e-10. Regulatory proteins were predicted using the Predicted Prokaryotic Regulatory Proteins web application\textsuperscript{36}. Detection of SNPs was performed using the Snippy software based on the alignment of unassembled reads with the \textit{P. aeruginosa} PAO1 genome sequence using default parameters\textsuperscript{7}. The hypothetical effect of non-synonymous variants were analyzed using the HOPE web application\textsuperscript{12}.

**Phylogenomic analysis.** Genome sequences of CCBH4851 and strains listed in Table 4 were used to perform whole-genome phylogeny analysis. The analysis was performed by CSI Phylogeny web application using default parameters\textsuperscript{31}. A core genome phylogeny tree was inferred and edited with FigTree (available at \url{http://tree.bio.ed.ac.uk/software/figtree/}).

**Antimicrobial susceptibility testing.** Antimicrobial susceptibility testing was performed according to the EUCAST guidelines using the broth microdilution method\textsuperscript{31}. Four major antimicrobial classes were assessed: aminoglycosides (gentamicin), quinolones (ciprofloxacin), beta-lactams (imipenem) and lipopeptides (polymyxin E or colistin). The MIC was determined as the lowest concentration of antimicrobial agent that prevented visible growth of the microorganism. Classification as susceptible, intermediate or resistant was based on EUCAST criteria\textsuperscript{39}. Briefly, resistance breakpoints for ciprofloxacin, imipenem and colistin are \( >0.5 \) \( \mu \)g/mL, \( >4 \) \( \mu \)g/mL and \( >2 \) \( \mu \)g/mL, respectively. Gentamicin has an "IE" status, meaning there is no sufficient evidence that the organism is a good target for therapy with the agent, MIC may be reported but without an S, I or R assignment. For the broth microdilution method, a total of six biological replicates with two technical replicates each were performed divided in three independent experiments. With exception of colistin susceptibility testing, that showed growth in skipped wells with no pattern, the same visual reading was observed in all experiments.

**Motility assays.** Motility assays were performed as described previously with slight modifications\textsuperscript{39}. Luria–Bertani (LB) agar was used as growth medium. Swimming plates (0.3% agar) were inoculated with a sterile toothpick and incubated for 16 h at 37\textdegree C. Swarming plates (0.6% agar) were inoculated with a sterile toothpick and the plates were incubated for 24 h at 37\textdegree C. For twitching, cells were inoculated with a toothpick through a LB layer (1% agar) to the bottom of the petri dish. Plates were incubated for 24 h at 37\textdegree C. After incubation, the agar was discarded, the plates were washed to remove unattached cells, and stained with crystal violet (1% [wt/vol] solution). All motilities were measured as the circular zone formed by bacterial migration away from the inoculation point. Additionally, the stain in the twitching plates was solubilized in ethanol (95% [vol/vol] solution) and the absorbance was measured at 570 nm using an ethanol solution as blank. With exception of the stain quantification in the twitching plates, at least three biological replicates with three technical replicates each were performed to three independent experiments. For stain quantification in the twitching plates, a total of five biological replicates with two technical replicates each were performed divided in two independent experiments.

**Biofilm formation assay.** The ability of \textit{P. aeruginosa} to form biofilms was assessed as described previously\textsuperscript{37}. Briefly, cells were inoculated in 96-well plates containing LB or Mueller–Hinton medium. Plates were incubated for 20–24 h at 37\textdegree C. After incubation, the medium was discarded, the plates were washed with water to remove unattached cells and stained with crystal violet (1% [wt/vol] solution). Then, the crystal violet was solubilized in ethanol (95% [vol/vol] solution), and the absorbance was measured at 570 nm using an ethanol solution as blank. \textit{P. aeruginosa} PAO1 was used as positive control. A total of seven biological replicates with at least three technical replicates each were performed divided in two independent experiments.

**Statistical analysis.** Statistics was performed in GraphPad Prism software. Student’s t test was used to test whether the results of phenotypic experiments were significantly different (\( P \) value < 0.05). Data were expressed as mean \( \pm \) standard error of the mean. In the figures, \( * \) equals \( P \) value < 0.05, \( ** \) equals \( P \) value < 0.01, \( *** \) equals \( P \) value < 0.001 and \( **** \) equals \( P \) value < 0.0001.

**Data availability** The genome sequence and annotation data generated in this study have been submitted to the GenBank database (\url{https://www.ncbi.nlm.nih.gov/genbank/}) under accession number CP021380.

Received: 13 March 2020; Accepted: 16 July 2020
Published online: 06 August 2020

**References**

1. Hong, D. J. et al. Epidemiology and characteristics of metallo-beta-lactamase-producing Pseudomonas aeruginosa. Infect. Chemother. 47, 81–97. \url{https://doi.org/10.3947/ic.2015.47.2.81} (2015).
2. Sawa, T., Kooguchi, K. & Moriyama, K. Molecular diversity of extended-spectrum \( \beta \)-lactamases and carbapenemases, and antimicrobial resistance. J. Intensive Care 8, 13. \url{https://doi.org/10.1186/s40560-020-0429-6} (2020).
3. Tooke, C. L. et al. \( \beta \)-Lactamases and \( \beta \)-Lactamase Inhibitors in the 21st Century. J. Mol. Biol. 431, 3472–3500. \url{https://doi.org/10.1016/j.jmb.2019.04.002} (2019).
4. Agência Nacional de Vigilância Sanitária. Boletim segurança do paciente e qualidade em serviços de saúde no. 17: Avaliação dos indicadores nacionais das infecções relacionadas à assistência à saúde (IRAS) e resistência microbiana do ano de 2017 (2019).
5. Escandón-Vargas, K., Reyes, S., Gutiérrez, S. & Villegas, M. V. The epidemiology of carbapenemases in Latin America and the Caribbean. Expert Rev. Anti-Infective Ther. 15, 277–297. \url{https://doi.org/10.1080/14787210.2017.1268918} (2017).
6. Silveira, M., Albano, R., Asensi, M. & Assel, A. P. C. The draft genome sequence of multidrug-resistant Pseudomonas aeruginosa strain CCBH4851, a nosocomial isolate belonging to clone SP (ST277) that is prevalent in Brazil. Memórias do Inst. Oswaldo Cruz 109, 1086–1087. \url{https://doi.org/10.1590/0074-0276140336} (2014).
11. The European Committee on Antimicrobial Susceptibility Testing. EUCAST reading guide for broth microdilution (2019).
10. Roy, P. H. Pseudomonas genomes: diverse and adaptable. 
13. Shu, J.-C.
15. Poole, K. Aminoglycoside resistance in Pseudomonas aeruginosa.
16. Chung, J. C. S.
17. Mann, E. E. & Wozniak, D. J. Pseudomonas biofilm matrix composition and niche biology.
21. O’Toole, G. A. & Kolter, R. Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development.
25. Lewis, K. Persister cells.
24. Kung, V. L., Ozer, E. A. & Hauser, A. R. The accessory genome of Pseudomonas aeruginosa.
27. Siguier, P., Gourbeyre, E. & Chandler, M. Bacterial insertion sequences: their genomic impact and diversity.
33. Ramanathan, B.
28. van Belkum, A.
32. López-Causapé, C., Cabot, G., Del Barrio-Tofiño, E. & Oliver, A. The versatile mutational resistome of Pseudomonas aeruginosa.
31. Tian, Z.-X.
29. Gophna, U.
40. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. 
44, 563–567. https://doi.org/10.1038/nature10829 (2010).
35. Ramanathan, B. et al. Next generation sequencing reveals the antibiotic resistant variants in the genome of Pseudomonas aeruginosa. PLoS ONE 12, e0182524. https://doi.org/10.1371/journal.pone.0182524 (2017).
49. Rodriguez-Martinez, J.-M., Poirel, L. & Nordmann, P. Extended-spectrum cephalosporinases in Pseudomonas aeruginosa. Antimicrob. Agents Chemother. 53, 1766–1771. https://doi.org/10.1128/AAC.02047-09 (2009).
53. Zimin, A. V. et al. The MaSuRCA genome assembler. Bioinformatics. 29, 2669–2677. https://doi.org/10.1093/bioinformatics/bts499 (2013).
64, 7563–7568. https://doi.org/10.3947/ic.2015.47.2.816 (2015).
50. Höfler, M. et al. Diversity of virulence genes in multidrug resistant Pseudomonas aeruginosa isolated from burn wound infections. Microb. Pathog. 115, 251–256. https://doi.org/10.1016/j.micpat.2017.03.002 (2017).
51. Lewis, K. Persistor cells. Annu. Rev. Microbiol. 64, 357–372. https://doi.org/10.1146/annurev-microbi-012814-064156 (2010).
52. Liu, Y. et al. Structural insights into the transcriptional regulation of HigBA toxin-antitoxin system by antitoxin HigA in Pseudomonas aeruginosa. Front. Microbiol. 10, 3158. https://doi.org/10.3389/fmicb.2019.03158 (2019).
53. Singh, S., Gourbeyre, E. & Delcour, K. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol. Rev. 38, 865–891. https://doi.org/10.1007/s10010-014-0037-z (2014).
54. van Belkum, A. et al. Phylogenetic distribution of CRISPR-Cas systems in antibiotic-resistant Pseudomonas aeruginosa. mbio. e01796–01715. https://doi.org/10.1128/mBio.01796-15 (2015).
55. Gophna, U. et al. No evidence of inhibition of horizontal gene transfer by CRISPR-Cas on evolutionary timescales. ISME J. 9, 2021. https://doi.org/10.1038/s41373-020-0094-9 (2021).
56. Maseda, H., Saito, K., Nakajima, A. & Nakae, T. Variation of the mexT gene, a regulator of the MexEF-oprN efflux pump expression in wild-type strains of Pseudomonas aeruginosa. FEMS Microb. Lett. 192, 107–112. https://doi.org/10.1007/s10010-014-0429-6 (2013).
57. Tian, Z.-X. et al. MexT modulates virulence determinants in Pseudomonas aeruginosa independent of the MexEF-OprN efflux pump. Microb. Pathog. 47, 237–241. https://doi.org/10.1016/j.micpat.2009.02.009 (2009).
58. López-Queciaspe, C., Cabot, G., Del Barrio-Tofiño, E. & Oliver, A. The accessory genome of Pseudomonas aeruginosa. Front. Microbiol. 9, 685. https://doi.org/10.3389/fmicb.2018.00685 (2018).
59. Ramanathan, B. et al. Next generation sequencing reveals the antibiotic resistant variants in the genome of Pseudomonas aeruginosa. PLoS ONE 12, e0182524. https://doi.org/10.1371/journal.pone.0182524 (2017).
60. Rodriguez-Martinez, J.-M., Poirel, L. & Nordmann, P. Extended-spectrum cephalosporinases in Pseudomonas aeruginosa. Antimicrob. Agents Chemother. 53, 1766–1771. https://doi.org/10.1128/AAC.02047-09 (2009).
61. Zimin, A. V. et al. The MaSuRCA genome assembler. Bioinformatics. 29, 2669–2677. https://doi.org/10.1093/bioinformatics/bts499 (2013).
62. Otto, T. D., Dillon, G. P., Degrave, W. S. & Berriman, M. RATT: rapid annotation transfer tool. Nucleic Acids Res. 39, e57. https://doi.org/10.1093/nar/gkw670 (2011).
63. Besemer, J., Lomsadze, A. & Borodovsky, M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29, 2607–2618 (2001).
64. Lagesen, K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35, 3001–3008. https://doi.org/10.1093/nar/gkm160 (2007).
65. Lowe, T. M. & Chan, P. P. iRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44, W54–W57. https://doi.org/10.1093/nar/gkw113 (2016).
66. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35, D61–D65. https://doi.org/10.1093/nar/gkq482 (2007).
67. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28, 33–36 (2000).
68. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44, D457–D462. https://doi.org/10.1093/nar/gkv1070 (2016).
69. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45, D158–D169. https://doi.org/10.1093/nar/gkw1099 (2017).
70. Mitchell, A. L. et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 47, D351–D360. https://doi.org/10.1093/nar/gky1100 (2019).
71. Claudel-Condé, C., Chevalet, C., Faraut, T. & Kahn, D. Enzyme-specific profiles for genome annotation: PRIAM. Nucleic Acids Res. 31, 6633–6639 (2003).
72. Rutherford, K. et al. Artemis: sequence visualization and annotation. Bioinformatics (Oxford, England) 16, 944–945 (2000).
47. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645. https://doi.org/10.1101/gr.092759.109 (2009).

48. Overbeek, R., Fonstein, M., D’Souza, M., Pusch, G. D. & Maltsev, N. The use of gene clusters to infer functional coupling. Proc. Natl. Acad. Sci. 96, 2896–2901. https://doi.org/10.1073/pnas.96.6.2896 (1999).

49. Medeiros Filho, F., do Nascimento, A. P. B., dos Santos, M. T., Carvalho-Assef, A. P. D. & da Silva, F. A. B. Gene regulatory network inference and analysis of multidrug-resistant Pseudomonas aeruginosa. Memrias Do Instituto Oswaldo Cruz 114, e190105. https://doi.org/10.1590/0071-02760190105 (2019).

50. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggng-nmapper. Mol. Biol. Evol. 34, 2115–2122. https://doi.org/10.1093/molbev/msx148 (2017).

51. Bertelli, C. et al. Islandviewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 45, W30–W35. https://doi.org/10.1093/nar/gkx343 (2017).

52. Varani, A. M., Siguier, P., Gourbeyre, E., Charneau, V. & Chandler, M. Iggasa is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol. 12, R30. https://doi.org/10.1186/gb-2011-12-3-r30 (2011).

53. Couvin, D. et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for cas proteins. Nucleic Acids Res. 46, W246–W251. https://doi.org/10.1093/nar/gkj425 (2018).

54. Ia, B. et al. CARD 2017: expansion and model-centric curating of the comprehensive antibiotic resistance database. Nucleic Acids Res. 45, D566–D573. https://doi.org/10.1093/nar/gkw1004 (2017).

55. Liu, B., Zheng, D., Jin, Q., Chen, L. & Yang, J. VFDX 2019: a comparative pathogenic genome platform with an interactive web interface. Nucleic Acids Res. 47, D687–D692. https://doi.org/10.1093/nar/gky1080 (2019).

56. Barakat, M., Ortel, P. & Whitworth, D. E. P2RP: a web-based framework for the identification and analysis of regulatory proteins in prokaryotic genomes. BMC Genom. 14, 269. https://doi.org/10.1186/1471-2164-14-269 (2013).

57. Venselaar, H., Te Beek, T. A. H., Kuipers, R. K. P., Hekkelman, M. L. & Vriend, G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinform. 11, 548. https://doi.org/10.1186/1471-2105-11-548 (2010).

58. Kaas, R. S., Leekitcharoenphon, P., Aarestrup, F. M. & Lund, O. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS ONE 9, e104984. https://doi.org/10.1371/journal.pone.0104984 (2014).

59. The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Clin. Microbiol. Infect. 18, 12–24. https://doi.org/10.1111/imi.12378 (2012).

60. Déziel, E., Comeau, Y. & Villemur, R. Initiation of biofilm formation by Pseudomonas aeruginosa 57rp correlates with emergence of hyperpiliated and highly adherent phenotypic variants deficient in swimming, swarming, and twitching motilities. J. Bacteriol. 183, 1195–1204. https://doi.org/10.1128/JB.183.4.1195-1204.2001 (2001).

61. O’Toole, G. A. Microtiter dish biofilm formation assay. Journal of Visualized Experiments: JoVE. https://doi.org/10.3791/2437 (2011).

62. Lee, J.-Y. & Ko, K. S. Mutations and expression of PmrAB and PhoPQ related with colistin resistance in Pseudomonas aeruginosa clinical isolates. Diagn. Microbiol. Infect. Dis. 78, 271–276. https://doi.org/10.1016/j.diagmicrobio.2013.11.027 (2014).

Acknowledgements

This work was supported by Programa Fiocruz de Fomento à Inovação - Inova Fiocruz (VPPCB-007-FIO-18-2-29) and Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (E-26/202.554/2019). This study was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. Hermes Senger is supported by Fundação de Apoio à Pesquisa do Estado de São Paulo (2018/00452-2).

Author Contributions

A.N. designed the study, performed computational analyses and biological experiments, and wrote the manuscript. H.P. performed biological experiments. F.F., H.S. and R.A. provided computational support. L.A., M.S., A.C., F.S. supervised the study. All authors provided critical feedback and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-69944-6.

Correspondence

and requests for materials should be addressed to A.P.B.N. or F.A.B.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020