Comparative genomic analysis of gene clusters of *Pseudomonas aeruginosa* that define specific biofilm formation in deciphering target regions for novel treatment options

**CURRENT STATUS:** UNDER REVIEW

Michael Ambutsi  ambutsim@gmail.com
Masinde Muliro University of Science and Technology Faculty of Science
*Corresponding Author*
ORCiD: 0000-0001-8841-634X

Oleg Reva
University of Pretoria

Patrick Okoth
Masinde Muliro University of Science and Technology

**DOI:**
10.21203/rs.2.21619/v1

**SUBJECT AREAS**
*Applied & Industrial Microbiology*  *General Microbiology*

**KEYWORDS**
*Biofilm formation, antibiotic resistance, comparative genomics, Pseudomonas aeruginosa, biomarkers*
Abstract

Background

*Pseudomonas aeruginosa* is an opportunistic pathogen associated with numerous nosocomial infections that are difficult to treat as a result of natural resistance to various antibiotics, particularly because of biofilm formation. The purpose of this study was to determine the distribution of biofilm formation genes in sequences of this opportunistic pathogen and their association with different ecological niches. In total, 13 genes responsible for biofilm formation by *P. aeruginosa* were identified and used in the study. They were clustered into seven categories based on the role they play in the biofilm formation process. The study also analyzed 185 complete genome sequences of *P. aeruginosa* strains retrieved from the NCBI and IPCD databases. These were classified into 14 categories based on the ecological niches they occupy.

Results

Phylogenetic analyses of the biofilm formation genes indicated a strong co-evolution of a majority of these genes, n=10 . Exceptions were the genes fliC, algD, and algU which may have been exchanged by horizontal gene transfer or evolved faster than the other genes of this functional group as they are more important in terms of a proper response of the biofilm formation to specific environmental stimuli in different habitats. The BLAST Ring Image Generator (BRIG) analysis was used to visualize the distribution of biofilm formation genes between different strains of *P. aeruginosa*.

Conclusions

fliC, algD, and algU genes were identified as potential targets for antibiofilm therapies. These findings could inform the development of antibiofilm therapies that target processes mediated by these genes. Also, this study provides useful information that can guide the direction of future research.
Background

Pseudomonas aeruginosa accounts for 11% of nosocomial infections with immunocompromised patients being the most susceptible (1). With a mortality rate of 60%, the opportunistic pathogen ranks high among the Gram-negative bacterial killers (2). The high mortality rate is attributed to antibiotic resistance that often is a result of mutations in chromosomal genes, horizontal gene transfer and acquisition of virulence plasmids bearing genes encoding for beta-lactamases and other drug resistance genetic determinants (3, 4). The antibiotic resistance is compounded by biofilms that cannot be easily destroyed (5). Pathogens often resort to the formation of biofilms to protect themselves from innate immune functions (6). The microorganism’s susceptibility to mutagenesis is also increased when it experiences an increase in biofilm formation. This is likely to result in less sensitivity to antibiotics further fueling the antibiotic resistance ability of the organism (7). It is important that the gene clusters responsible for biofilm formation in P. aeruginosa are identified so that novel therapies can be centered on controlling this important phenomenon. It may be also of importance to be able to identify virulence genes in metagenomic DNA reads to monitor the distribution of nosocomial infections.

Pseudomonas is known to continually sense the environment and adapt to different phenotypic forms in a bid to increase its chances of survival within the human host. When the immune system is compromised, the microorganism steps up its pathogenesis as is the case in immunocompromised patients (6). It has also been previously reported that biofilm cells can modulate some of their genes in a bid to quickly adapt to harsh environmental conditions (8). These cells can either opt to decrease the permeability of the cell wall, express multidrug efflux pumps at high levels or activate antibiotic modifying enzymes as they try to evade therapeutic efforts castigated against them.
Previous studies have suggested that development of novel therapies for refractory P. aeruginosa may be augmented by understanding the mechanisms of biofilm-specific adaptations exhibited by the bacteria (8). The expression of drug resistance genes in P. aeruginosa often accompanies biofilm formation. This feature provides an evolutionary advantage to the bacteria allowing it to evade the effects of drugs as well as the defense system of the host organism (8). Apart from this synergy, previous studies have also revealed that the extracellular matrix of bacterial biofilms can act as a diffusion barrier for antibiotics (9). There is a clear need for a better understanding of the interplay between the genes involved in biofilm formation by P. aeruginosa that gives more credence to this study. Such knowledge will reveal potential molecular targets for development of novel effective therapies.

Biofilm formation in P. aeruginosa relies on a natural heterogeneity of bacterial populations switching to biofilm formation when the nutrients become scarce and waste products accumulated (10). Several antimicrobial agents have previously been developed to target these population development phase (11). Pathogens have adapted to these therapeutic agents by either population behavior changes or eliminating molecular targets (12). Phase variations in bacterial populations are mediated by a number of biofilm formation genes including amrZ, which modulates the psl operon responsible for microcolony formation (10). Our study sought to decipher the roles of various biofilm formation genes and tried to relate the functional properties with the evolutionary relationships and location of these genes in the genomes of the pathogen. Although the biofilm formation is an important factor in the development of antibiotic resistance, little is known about the conservation and variation patterns of the genes responsible for this attribute in P. aeruginosa. Rather than following a systematic exploration, new findings around biofilm formation in the opportunistic pathogen have been based on chance (13).
This study sought to employ various comparative genomics tools to establish the conservation and variation patterns of biofilm formation genes among different strains of *P. aeruginosa*.

*P. aeruginosa* consists of different strains that occupy different niches within their human hosts and are classified under three clades based on their evolutionary relationships. The evolutionary differences have previously been linked to biotope associations within the host and environmental niches (14). Previous studies on the opportunistic pathogen are yet to provide reliable explanations for the pathogen’s population structure and exhaustive analyses of the NCBI repositories remain scanty and disjointed (15). The National Center for Biotechnology Information (NCBI) contains approximately 176 complete genomes of *Pseudomonas aeruginosa* strains isolated from different ecological niches. These sequences provide a valuable source for investigating the complexity and diversity of the various strains of the opportunistic pathogen (16). Given the large number of genomes present in the NCBI GenBank database, an in silico approach to characterize the genes involved in biofilm formation in *P. aeruginosa* looks feasible.

Sequences of genes responsible for biofilm formation in the opportunistic pathogen were retrieved from the NCBI GenBank and analyzed using different comparative genomics tools in a bid to characterize these genes in various strains of *P. aeruginosa*. The study identified conservation patterns of genes in the opportunistic pathogen. These were designated as regions of interest and identified as biomarkers for novel treatment options. Comparative genomics was used to identify variations in these genes especially in flexible regions of *P. aeruginosa* genomes. The findings of the study revealed potential biomarkers for novel treatment options and shed more light on the survival mechanism of *P. aeruginosa* within the human host.

**Results**
**P. aeruginosa** Sequence Retrieval

The study sought to retrieve complete genome sequences of *P. aeruginosa* strains isolated from various ecological niches. These sequences were analyzed by different comparative genomics tools in a bid to characterize the biofilm formation genes with respect to different strains of the pathogen. A total of 185 complete genome sequences of *P. aeruginosa* strains from the NCBI and IPCD databases were retrieved for analysis as indicated in Table 1. Microorganisms were classified into 14 categories based on the ecological niches the different strains were isolated from. The study used metadata on each strain to complete this classification. Table 1 shows niche-specific categories of *P. aeruginosa* isolates which were used for downstream analyses.

| Ecological niche      | Analyzed sequences |
|-----------------------|--------------------|
| Abscess               | 2                  |
| Blood                 | 16                 |
| Bronchial             | 6                  |
| Cell culture          | 4                  |
| Clinical              | 10                 |
| Dental                | 1                  |
| Environment           | 8                  |
| Eye                   | 2                  |
| Lung                  | 1                  |
| Sputum                | 26                 |
| Trachea aspirates     | 5                  |
| Urine                 | 7                  |
| Wound                 | 9                  |
| Unclassified          | 98                 |
| **Total**             | **185**            |

Figure 1 indicates the niche associations of the strains selected from NCBI. The selection doesn’t represent any real prevalence of *P. aeruginosa* in nature. It is biased by how strains are selected for sequencing.

**Identification of Biofilm Formation Genes**

Using genome annotation data and sequence alignments, 13 biofilm formation genes common in most of the strains of the *P. aeruginosa* were identified and grouped into corresponding COGs represented by individual FASTA files as indicated in table 2. Each file
contained 44 protein sequences of the respective genes selected from every *P. aeruginosa* reference genome. Table 3 indicates the names, GenBank accession number, size and number of biofilm associated genes represented by individual COGs.

| ID      | Gene name | Accession Number |
|---------|-----------|------------------|
| 882251  | pslJ      | NC_002516.2      |
| 879704  | pslG      | NC_002516.2      |
| 878490  | pslE      | NC_002516.2      |
| 882052  | fliC      | NC_002516.2      |
| 882125  | algU      | NC_002516.2      |
| 879406  | algC      | NC_002516.2      |
| 881782  | rsaL      | NC_002516.2      |
| 879474  | gshA      | NC_002516.2      |
| 879004  | algD      | NC_002516.2      |
| 882372  | ppyR      | NC_002516.2      |
| 879143  | arnB      | NC_002516.2      |
| 8819333 | htpG      | NC_002516.2      |
| 878223  | gshB      | NC_002516.2      |

**Table 3**

The names, GenBank accession numbers, size, number of annotated genes and ecological niches of the 44 Pseudomonas aeruginosa genomes
| Scientific Names         | GenBank Accession Number | Size (kbps) | Number of annotated genes | GC (%) | Ecological niche          |
|--------------------------|--------------------------|-------------|----------------------------|--------|---------------------------|
| P. a PA01                | NC_002516                | 6,264.404   | 5700                       | 66.56  | Unclassified              |
| P. a strain 24Pae112     | NZ_CP029605              | 7097.241    | 6596                       | 65.99  |                          |
| P. a strain 268          | NZ_CP032761              | 7030.474    | 6604                       | 65.91  |                          |
| P. a strain B17932       | NZ_CP034436              | 6744.658    | 5943                       | 65.94  |                          |
| P. a strain BA15561      | NZ_CP033432              | 6793.961    | 5813                       | 65.84  |                          |
| P. a strain NCTC 12903   | NZ_LR134309              | 6839.985    | 6431                       | 66.09  | Blood                    |
| P. a strain PA1207       | NZ_CP022001              | 7411.863    | 6813                       | 65.70  | Bronchial                 |
| P. a strain PA1242       | NZ_CP022002              | 7050.510    | 6303                       | 65.80  |                          |
| P. a strain PABL012      | NZ_CP031659              | 6546.467    | 6089                       | 66.29  |                          |
| P. a strain PABL017      | NZ_CP031660              | 6503.460    | 6019                       | 66.31  |                          |
| P. a strain Pa58         | NZ_CP021775              | 7241.575    | 6673                       | 65.80  |                          |
| P. a strain Pa84         | NZ_CP021999              | 6566.724    | 6058                       | 66.23  |                          |
| P. a strain Pa124        | NZ_CP021774              | 7008.516    | 6479                       | 65.84  |                          |
| P. a strain Pa127        | NZ_CP022000              | 7148.302    | 6565                       | 65.74  |                          |
| P. a strain GIMC5015:PAKB6| NZ_CP034429             | 6258.491    | 5772                       | 66.53  |                          |
| P. a strain H26023       | NZ_CP033685              | 6729.216    | 6260                       | 66.21  |                          |
| P. a strain NCTC11445    | NZ_LR134308              | 6766.292    | 6378                       | 66.06  |                          |
| P. a paerg002            | NZ_LR130527              | 6451.470    | 5935                       | 66.40  |                          |
| P. a paerg003            | NZ_LR130530              | 6433.962    | 5945                       | 66.40  |                          |
| P. a paerg004            | NZ_LR130531              | 6452.809    | 5936                       | 66.40  |                          |
| P. a paerg005            | NZ_LR130534              | 6931.425    | 6427                       | 66.00  | Clinical                 |
| P. a paerg009            | NZ_LR130533              | 6941.287    | 6352                       | 65.98  |                          |
| P. a paerg010            | NZ_LR130536              | 6433.960    | 5950                       | 66.40  |                          |
| P. a paerg011            | NZ_LR130535              | 6434.133    | 5946                       | 66.40  |                          |
| P. a paerg012            | NZ_LR130537              | 6434.020    | 5948                       | 66.40  |                          |
| P. a strain L10          | NZ_CP019338              | 6661.962    | 6119                       | 66.13  | Environment              |
| P. a strain PA34         | NZ_CP032552              | 6810.079    | 6314                       | 66.07  | Eye                       |
| P. a C-NN2 isolate       | NZ_LT883143              | 6902.967    | 6412                       | 66.12  | Lung                      |
| P. a strain H25883       | NZ_CP033686              | 6706.800    | 6236                       | 66.15  |                          |
| P. a strain H26027       | NZ_CP033684              | 7079.598    | 6650                       | 66.07  |                          |
| P. a strain MRSN12280    | NZ_CP028162              | 7070.928    | 6597                       | 66.02  | Wound                    |
| P. a PAO1161             | NZ_CP032126              | 6383.803    | 5918                       | 66.42  |                          |
| P. a strain NCTC13715    | NZ_LR134330              | 6765.311    | 6288                       | 66.12  | Urine                    |
| P. a strain FDAARGOS 505  | NZ_CP033832              | 7029.824    | 6520                       | 65.87  | Trachea                   |
| P. a strain AES1M        | NZ_CP037925              | 6373.893    | 5833                       | 66.48  |                          |
| P. a strain AES1R        | NZ_CP037926              | 6373.893    | 5833                       | 66.48  |                          |
| P. a strain CCGU 70744   | NZ_CP023255              | 6859.232    | 6422                       | 66.04  |                          |
| P. a strain LW           | NZ_CP022478              | 6824.837    | 6271                       | 65.97  |                          |
| P. a strain PASGNDM345   | NZ_CP020703              | 6893.164    | 6432                       | 66.07  | Sputum                   |
| P. a strain PASGNDM699   | NZ_CP020704              | 6985.102    | 6545                       | 66.00  |                          |
| P. a strain SP2230       | NZ_CP034434              | 6976.603    | 6067                       | 65.74  |                          |
| P. a strain SP4527       | NZ_CP034409              | 7005.215    | 6123                       | 65.79  |                          |
| P. a strain SP4528       | NZ_CP033439              | 6877.287    | 6082                       | 65.85  |                          |
| P. a strain Y31          | NZ_CP030910              | 6831.076    | 6322                       | 66.15  |                          |

The retrieved sequences were further classified into four categories based on the role they play in the biofilm formation process. An additional category was created to cater for genes whose functions are not clearly understood yet. Specific roles were identified using related publications and metadata provided for each gene. Based on this, the genes were assigned to different classes as shown in Table 4.
### Comparative Genomic Analyses

From the comparison of the 13 COG-based ML phylogenetic trees, the study created a tree of relationships between different biofilm related genes using the treedist distance matrix (Figure 2). The phylogenetic analyses revealed four clusters. From the 13 biofilm formation genes analyzed, 10 genes fell into a single cluster. The \(algD\) and \(algU\) genes diverged the most from the other biofilm formation genes, while \(fliC\) was not completely divergent from the other genes which seem to have co-evolved together. While the study assumed that all the biofilm formation genes co-evolved together given that they belong to a group of functionally related genes that generally was confirmed by the obvious co-evolution of these genes – the divergence of the three genes may result from a horizontal gene transfer. Alternatively, it may be assumed that these genes evolved faster than other genes as they are more important in terms of a proper response of the biofilm formation to specific environmental stimuli in different habitats. It makes these genes potential targets for antibiofilm therapies. Both the \(algD\) and \(algU\) genes were classified as regulatory genes responsible for the regulatory stage of the biofilm formation process.

### Genome Mapping

To further investigate the role of horizontal gene transfer in the divergence observed in \(algU\), \(algD\) and \(fliC\), the study conducted a BRIG analysis of different strains of \(P. aeruginosa\). This analysis helped to determine the variable and conserved regions within the genome as well as the distribution of the biofilm formation genes. It also pointed

---

**Table 4**

| Classes       | No of Sequences | Genes                  |
|---------------|-----------------|------------------------|
| Adhesins      | 1               | \(ppyR\) (\(psl\))    |
| Repressors    | 1               | \(GshB\)               |
| Regulatory    | 5               | \(algC, algD, algU, amB, rsaL\) |
| Motility      | 3               | \(fliC, gshA, htpG\)   |
| Unclassified  | 3               | \(pslJ, pslE, pslG\),  |
| **Total**     | **13**          |                        |
further light to the meaning of these mutations in different ecological niches occupied by the ubiquitous pathogen. From the BRIG analysis the study created 14 gene maps, each map representing strains of the pathogen occupying different ecological niches. (Figure 3A to figure 3N). Figure 3A indicated the distribution of the biofilm formation genes among the sequences of the strains isolated from the 13 ecological niches that were identified by this study. Biofilm formation genes retrieved by the custom python scripts were highlighted in green. This analysis revealed that the pslE, pslG and pslJ genes were all located around the same locus (2450kbps) in a variable region of the genome of strains of *P. aeruginosa*. This gives credence to the coevolution between these set of genes. The three genes are part of the eleven psl genes required for Psl production and surface attachment which is necessary for biofilm formation in nonmucoid strains of *Pseudomonas aeruginosa* which don’t depend on alignate as the main biofilm polysaccharide (17). These genes are co-transcribed and our evolutionary analysis indicated that they co-evolve with each other. The other ten genes were located in different loci, all of which were conserved regions of the genome. Figure 3B to figure 3N indicated the distribution of the genes among sequences of strains isolated from each individual ecological niche. Besides the distribution of biofilm formation genes, the BRIG analysis was also used to identify the conserved and variable regions of the genome of the pathogen. This was done to further elucidate the differences in the distribution of genes in strains from different ecological niches.

**Discussions**

Protein AlgU has been identified as a founding member of the ExtraCytoplasmic Function (ECF) family - a group of products responsible for transcriptional regulation and response to environmental stress (18). This sigma factor has been associated with regulation of
genes having the AlgD’s promoter (19). Previous studies indicate that mutations in algU can affect mucoidy in P. aeruginosa or lead to a partially active AlgU (20, 21, 22). A study by Scalab et al. (2015), also indicated that this gene is important for P. aeruginosa to withstand various treatments hence the need for an adjustment in its activity by specific mutations (23). It may explain a rather specific evolution pattern displayed by this protein. Divergence of this gene as seen in Fig. 2 may result from selective accumulation of such mutations. The mucoidy phenotype improves the pathogenesis of P. aeruginosa infections as the pathogen acquires increased resistance to antibiotics and phagocytic killing while allowing it to evade the host’s immune response (24, 25). This prominent role of the gene makes it a suitable target for antibiofilm therapy. Further studies should be conducted to understand the exact mutations of this sigma factor and improve efforts to tackle antibiotic resistance witnessed in infections caused by the ubiquitous pathogens. On the other hand, algD is part of an operon that codes for proteins responsible for the biosynthesis, modification, and export of alignate, an important component of an exopolysaccharide matrix produced by chronic P. aeruginosa. Alignate helps the pathogen acquire resistance to antibiotics (tobramycin) while resisting phagocytosis and opsonization (26). The phylogenetic results are consistent with previous studies which indicated that alignate overproduction can be regulated at both the post-translational and genetic mutations (27). The divergence of the algD gene from the other biofilm formation genes indicates an adjustment of its activity to specific habitat needs via genetic mutations. Further studies around this gene could be performed to identify the specific mutation along with the consequences of mutations in this specific gene. A clear understanding of such mutations could come in handy in tackling the menace of antibiotic resistance exhibited by pathogenic strains of P. aeruginosa.

The gene fliC codes for the B-type flagellin protein that is necessary for biofilm formation
in *P. aeruginosa*. Previous studies have indicated that fliC mutant strains are hampered in their ability to cause diseases in model animals (28). Similar studies have shown that pilli expression does not contribute significantly to virulence in the absence of flagella. The expression of these gene is however downregulated in strains from cystic fibrosis isolates, allowing the pathogen to colonize the lungs (29). With the significant diversion from the other biofilm formation genes, we speculate that the fliC gene undergoes necessary mutations to enable the pathogenesis of different strains of *P. aeruginosa* and survival within different ecological niches.

This study aimed to identify biofilm formation genes, classify them based on the role they play in the biofilm formation process and analyze the distribution of these genes in the genomes of various strains of *P. aeruginosa*. All these objectives were aimed at identifying genes and processes that could serve as potential targets for novel antibiofilm therapies. Proper understanding of interactions mediated by these genes could potentially result in development of therapeutic agents that impair biofilm formation. Infections that are refractory to antibiotic treatments could be kept in check.

**Conclusion**

The diverse distribution of genes involved in biofilm formation in *P. aeruginosa* highlights the diversity of pathways the bacterium can explore in different ecological niches. The conservation and variability of some of these genes in particular niches offers the scientific body a lot to think about especially on the possibility of new antibiofilm therapies. It is possible that these evolutionary differences are associated with horizontal gene transfer, a phenomenon that could be the center of future investigations on biofilm formation genes. The study identified potential target genes for novel antibiofilm therapies, including fliC, algD and algU. The study’s hypothesis that gene clusters with similar functions evolved similarly was confirmed by the coevolution of these genes.
Divergence witnessed in three genes was associated with horizontal gene transfer or faster evolution of these genes. Further studies on the evolutionary relationships between these genes need to be done to identify the underlying role of the differences that we witnessed. Our study revealed that the biofilm formation process is dependent on the synergy between different classes of genes that modulate various steps in the overall process.

Methods

**P. aeruginosa Sequence Retrieval**

The study collated and analyzed whole genomes of *P. aeruginosa* to determine the conservation and variation patterns of biofilm formation genes in selected strains of *P. aeruginosa* strains. Complete genomes were downloaded from NCBI and the International *Pseudomonas* Consortium Database (https://ipcd.ibis.ulaval.ca/) as GBK files for complete genomes and Fasta files for individual genes. The study only selected “the complete genomes” and sequences available by December 2018 from these databases. The genome sequences were then categorized based on the ecological niches that they occupy within the human host based on the metadata provided by the databases.

**Identification of Biofilm Formation Genes**

For a comprehensive analysis of the phylogenetic relationship between the biofilm formation genes, the study sought to collect orthologous genes from representative genomes of *P. aeruginosa* based on amino acid sequence similarity using a reciprocal BLASTP search. Genes from different genomes showing at least 90% protein sequence similarity or above were grouped into clusters of orthologous genes (COG). An in-house Python 2.7 script was written to facilitate the retrieval of the biofilm formation gene sequences from the annotated genomes (GenBank files) of *P. aeruginosa*. The script was designed to create FASTA files for every COG, containing sequences of the respective
genes selected from all reference genomes.

**Comparative Genomic Analyses**

To infer phylogenetic relationships between sequenced *P. aeruginosa*, all proteins of every COG were aligned by MUSCLE algorithm (30). Individual COG alignments were edited by the program Gblocks (31) and then concatenated into a super-alignment used for a phylogenetic inference by the Neighbour-Joining (NJ) algorithm implemented in the program of PHYLIP 3.69 (Phylogeny Inference Package) (32).

For every COG of biofilm formation genes found in selected reference different strains, \( n = 13 \), the study conducted multiple sequence alignments using the MUSCLE algorithm (30). The default values of the MUSCLE algorithm were used. From the aligned sequences, maximum likelihood (ML) phylogenetic trees were created for every COG. The resulted phylogenetic trees were compared by the program treedist of PHYLIP 3.69 using the Branch Score Distance algorithm (32). An in-house Python script was used to reformat the treedist text output file into a matrix of distances between COG-based ML phylogenetic trees in PHYLIP format suitable for clustering of the trees by the NJ algorithm. The resulted dendogram was used to analyze grouping of co-evolved biofilm formation genes. Clustering of several genes together would mean their co-evolution while separating genes to different clusters would mean horizontal gene transfer exchange.

**Genome Mapping**

The BLAST Ring Image Generator (BRIG) was used to analyze different strains of *P. aeruginosa* to visualize the core and flexible genomes against a reference genome (33). The PAO1 strain was used as the reference genome for these analyses. The BRIG analyses were also performed to determine the distribution of the biofilm formation genes among the different strains of *P. aeruginosa* with the PAO1 used as the reference genome.

**Statistics**
Gene orthology was predicted by a reciprocal BLASTP alignment of protein sequences of all protein coding genes in one genome against all protein coding genes in another genome in a direct and reverse order. Those genes which showed the best BLASTP alignment hits reciprocally as query and subject sequences with 90% protein sequence identity or above and 75% coverage were considered as pairs of orthologous genes. Robustness of clustering of co-evolved genes was controlled by the bootstrapping analysis implemented in the program bootstrap of the package PHYLIP 3.69 using 100 replicates of the distance matrix.

Abbreviations

**BLAST**: Basic Local Alignment Search Tool

**BRIG**: BLAST Ring Image Generator

**COG**: Clusters of Orthologous Genes

**ECF**: ExtraCytoplasmic Function

**GBK**: GenBank

**IPCD**: International Pseudomonas Consortium Database

**ML**: Maximum Likelihood

**NCBI**: National Center for Biotechnology Information

**NJ**: Neighbor Joining

Declarations

**Ethics approval and consent to participate**

Not applicable

**Consent for publication**

Not applicable

**Availability of data and material**
All the data sets used in the study are available with no supplementary material.

**Competing interests**

The authors declare no conflict of interest.

**Funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Author’s contributions**

All authors contributed to the conceptualization of the work and writing of the manuscript. MA and OR performed the comparative genomic analyses and the statistics. MA and PO performed genome mapping. All authors read, revised and approved the final manuscript.

**Acknowledgements**

I thank Prof. Oleg Reva and Dr. Okoth Patrick for their critical comments and suggestions for improving the manuscript. Special thanks are given to the Department of Biological Sciences, SONAS, MMUST.

**References**

1. Balasubramanian D, Mathee K (2009). Comparative transcriptome analyses of *Pseudomonas aeruginosa*. Hum Genomics 3:349-361. doi:10.1186/1479-7364-3-4-361

2. Kim Yj, Jun YH, Kim YR, Park KG, Park YJ, Kang JY and Kim SI (2014). Risk factors for mortality in patients with *Pseudomonas aeruginosa* bacteremia; retrospective study of impact of combination antimicrobial therapy. *BMC infectious diseases, 14*, doi:10.1186/1471-2334-14-161

3. Okesola AO, Oni AA (2012). Occurrence of Extended-Spectrum Beta-Lactamase-Producing *Pseudomonas aeruginosa* Strains in South-West Nigeria. *Research Journal of Medical Sciences*. 6 (3):93–96. doi:10.3923/rjmsci.2012.93.96
4. Araujo BF, Ferreira ML, Campos PA, Rover S, Batistao DW, Dantas RC, Goncalves IR, Faria AL, Brito CS, Yokosawa J, Gontijo-Filho PP, Ribas RM (2016). Clinical and molecular epidemiology of multidrug-resistant *aeruginosa* carrying *aac* (60)-*lb-cr*, *qnrS1* and *blaSPM* genes in Brazil. PLoS One: e0155914. doi:10.1371/journal.pone.0155914

5. Farshadzadeh Z, Khosravi AD, Alavi SM, Parhizgari N and Hoveizavi H. (2014). Spread of Extended-Spectrum Beta-Lactamase Genes of *blaOXA-10*, *bla-PER-1* and *blaCTX-M* in *Pseudomonas aeruginosa* Strains Isolated from Burn Patients1. Burns 40, 1575-1580. doi:10.1016/j.burns.2014.02.008

6. Rao J, DiGiandomenico A, Artamonov M, Leitinger N, Amin AR and Goldberg J B (2011). Host derived inflammatory phospholipids regulate rahU (PAO122) gene, protein, and biofilm formation in *Pseudomonas aeruginosa*. *Cellular immunology*, 270(2), 95-102. doi:10.1016/j.cellimm.2011.04.011

7. David M Livermore (2002). Multiple Mechanisms of Antimicrobial Resistance in *Pseudomonas aeruginosa*: Our Worst Nightmare? *Clinical Infectious Diseases*, 34(5), 634-640. doi:10.1086/338782

8. Wang F, He Q, Yin J, Xu S, Hu W and Gu L (2018). BrlR from *Pseudomonas aeruginosa* is a receptor for both cyclic di-GMP and pyocyanin. *Nature communication*, 9(1),2563. doi:10.1038/s41467-018-05004-y

9. Olsen, I (2015). Biofilm-specific antibiotic tolerance and resistance. Eur. J. Clin. Microbiol. 34, 877–886. doi:10.1007/s10096-015-2323-z

10. Jones CJ, Ryder CR, Mann EE and Wozniak DJ (2013). AmrZ modulates *Pseudomonas aeruginosa* biofilm architecture by directly repressing transcription of the *psl* J. Bacteriol 195:1637-1644. doi:10.1128/JB.02190-12

11. Banin E, Brady KM and Greenberg EP (2006). Chelator-induced dispersal and killing of
**Pseudomonas aeruginosa** cells in a biofilm. Appl. Environ. Microbiology. 72:2064-2069. doi:10.1128/AEM.72.3.2064-2069.2006

12. Pamp Sj, Gjermansen M, Johansen HK and Tolker-Nielsen T (2008). Tolerance to antimicrobial peptide colistin in Pseudomonas aeruginosa biofilms is linked to metabolically active cells and depends on the *pmr* and *mexAB-oprM* Mol Microbiol. 68:223-240. doi:10.1111/j.1365-2958.2008.06152.x

13. Francisco CE, Ismael HG, Jorge EI (2019). Search for Cry proteins expressed by *Bacillus* genomes, using hidden Markov model profiles. 3 Biotech 9:13. doi:10.1007/s13205-018-1533-3

14. Wielhmann L, Cramer N, Tummler B (2015). Habitat-associated skew of clone abundance in the *Pseudomonas aeruginosa* Environ Microbiol Rep. 7(6):955-960. doi:10.1111/1758-2229.12340

15. Freschi L, Jeukens J, Kukavica-Ibrulj I, Boyle B, Dupont MJ, Laroche J, Larose S, Maaroufi H, Fothergill JL, Moore M, Winsor GL, Aaron SD, Barbeau J, Bell SC, Burns JL, Camara M (2015). Clinical utilization of genomics data produced by the international *Pseudomonas aeruginosa* Front Microbiol 6:1036. doi:10.3389/fmicb.2015.01036

16. Bruggemann H, Leticia B, Romario O, Paula C, Souza AV, Jensen A, Poehlein A, Brzuszkiewicz E, Doi AM, Pasternak J, Martino MDV, Severino P (2018). Comparative Genomics of Non-outbreak Pseudomonas aeruginosa Strains Underlines Genome Plasticity and Geographic Relatedness of the Global Clone ST235. Genome Biol. Evol. 10(7):1852–1857. doi:10.1093/gbe/evy139

17. Bryd MS, Sadovskaya I, Vinogradov E, Lu H, Sprinkle AB, Richardson SH & Wozniak DJ (2009). Genetic and biochemical analyses of the *Pseudomonas aeruginosa* Psl exopolysaccharide reveal overlapping roles for polysaccharide synthesis enzymes in Psl and LPS production. *Molecular microbiology*, 73(4), 622-638. doi:10.1111/j.1365-
18. Sineva E, Savkina M, & Ades SE (2017). Themes and variations in gene regulation by extracytoplasmic function (ECF) sigma factors. *Current opinion in microbiology, 36*, 128-137. doi:10.1016/j.mi.2017.05.004.

19. Yin Y, Damron FH, Withers TR, Pritchett CL, Wang X, Schurr MR & Yu HD (2013). Expression of mucoid induction factor MucE is dependent upon the alternate sigma factor AlgU in *Pseudomonas aeruginosa*. *BMC Microbiol* **13**, 232. doi:10.1186/1471-2180-13-232

20. Schurr MJ, Martin DW, Mudd MH & Deretic V (1994). Gene cluster controlling conversion to alginic-acid production phenotype in *Pseudomonas aeruginosa*: functional analysis in a heterologous host and role in the instability of mucoidy. *Bacteriol* **176**(11):3375-3382. doi:10.1128/jb.176.11.3375-3382.1994

21. DeVries CA & Ohman DE (1994). Mucoid-to-nonmucoid conversion in alginic-acid producing *Pseudomonas aeruginosa* often results from spontaneous mutations in algT, encoding a putative alternate sigma factor, and shows evidence for autoregulation. *J Bacteriol* **176**(21):6677-6687. doi:10.1128/jb.176.21.6677-6687.1994

22. Damkiaer S, Yang L, Molin S & Jelsbak L (2013). Evolutionary remodeling of global regulatory networks during long-term bacterial adaptation to human hosts. *Proc Natl Acad Sci USA*. **110**(19):7766-7771. doi: 10.1073/pnas.1221466110

23. Scalan PD, Hall AR, Blackshields G, Friman VP, Davis MR, Goldberg JB & Buckling A (2015). Coevolution with bacteriophages drives genome-wide host evolution and constrains the acquisition of abiotic-beneficial mutations. *Molecular biology and evolution*, **32**(6), 1425-1435. doi:10.1093/molbev/msv032

24. Pier GB, Coleman F, Grout M, Franklin M & Ohman DE (2001). Role of alginic O
acetylation in resistance of mucoid Pseudomonas aeruginosa opsonic phagocytosis. 

Infect Immun 69(3):1895-1901. doi:10.1128/IAI.69.3.1895-1901.2001

25. Leid JG, Willson CJ, Shirtliff ME, Hassett DJ, Parsek MR & Jeffers AK (2005). The exopolysaccharide alignate protects Pseudomonas aeruginosa biofilm bacteria from IFN-gamma-mediated macrophage killing. Immunol 175(11): 7512-7518. doi: 10.4049/jimmunol.175.11.7512

26. Okkotsu Y, Little AS & Schurr MJ (2014). The Pseudomonas aeruginosa AlgZR two-component system coordinates multiple phenotypes. Frontiers in cellular and infection microbiology, 4, doi:10.3389(fcimb).2014.00082

27. Mena A, Smith EE, Burns JL, Speert DP, Moskowitz SM, Perez JL & Oliver A (2008). Genetic adaption of Pseudomonas aeruginosa to the airways of cystic fibrosis patients is catalyzed by hypermutation. Journal of bacteriology, 190(24), 7910-7917. doi:10.1128/JB.01147-08

28. Feldman M, Bryan R, Rajan S, Scheffler L, Brunnet S, Tang H & Prince A (1998). Role of flagella in pathogenesis of Pseudomonas aeruginosa pulmonary infection. Infection and immunity, 66(1), 43-51. doi: 10.1128/IAI.66.1.43-51.1998

29. Sriramulu DD, Nimtz M & Romling U (2005). Proteome analysis reveals adaptation of Pseudomonas aeruginosa to the cystic fibrosis lung environment. Proteomics 5(14), 3712-21. doi: 10.1002/pmc.200401227

30. Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), 1792-1797. doi:10.1093/nar/gkh340

31. Alikhan NF, Petty NK, Ben Zakour NL and Beatson SA (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12:402. doi:10.1186/1471-2164-12-402
Figures

Distribution of Pseudomonas aeruginosa strains in different ecological niches

Figure 1

Distribution of selected strains of Pseudomonas aeruginosa strains in the various ecological niches. Strains isolated from the sputum niche were the most abundant while the isolates from the lungs and dental niches were the least abundant.
Co-evolution analyses of genes responsible for biofilm formation in strains of P. aeruginosa. The NJ dendogram shows that the majority of gene COGs produced identical phylogenetic trees of the selected reference genomes that indicates a strong co-evolution of these genes. Exceptions were the genes fliC, algD and algU which may be exchanged by horizontal gene transfer or evolved faster than other
genes of this functional group. The tree is drawn to scale with branch lengths in the same units as those of the evolutionary distances used to infer the COG phylogenetic tree.

Figure 3

A. Visualization of genome comparison of the different strains of Pseudomonas aeruginosa classified based on the ecological niches they were sourced from. The BRIG analysis also included the distribution of biofilm formation genes within the genomes of different strains. Differences regarding the flexible genome can be seen. Different colors on the rings indicated significant matches while non-significant matches were represented by blanks. Biofilm formation genes retrieved by the python script were highlighted in green. B. BRIG analysis for bronchial sequences reveals that genes in the variable regions include PA4878 and lipC at locus 5400kbps, bfiS and bfiR at locus 4600kbps, amiC at locus
3800kbps and the pclass at locus 2400kbps. C. Brig analysis for clinical sequences indicated that genes in the variable regions included ppkA at locus 100kbp, p class at locus 2450kbps, amiC at locus 3750kbps, arnB and arnD at locus 3950kbps, wspC at locus 4100kbps, pA4625 at locus 5150kbos and algC at locus 6000kbps.D. BRIG analysis for cell culture sequences indicated the genes in the variable regions included the p class at locus 2450kbps, and the lipC and PA4781 at locus 5400kbps.E. BRIG analysis for dental sequences revealed that the genes in the variable regions included cupC1 at locus 1050kbps, fliC at locus 1200kbps, p class at locus 2450kbps, PA2572 at locus 3000kbps, PA2771 at locus 3150kbps, amiC at locus 3700kbps, PA3989 at locus 4500kbps, PA4108 at locus 4600kbps and bfiS and bfiR at locus 4700kbps.F. BRIG analysis for environmental sequences showed that genes in the variable regions included the p class at locus 2450kbps, PA2824 at locus 3150kbps, amiC at locus 3750kbps, morA at locus 5150kbps and PA4625 at locus 5200kbps. G. BRIG analysis for sequences from the eye indicated the genes in the variable regions included fliC at the 1050kbps locus, amiC at the locus 3750kbps, PA3989 at the locus 4450kbps and bfiS and bfiR at the locus 4700kbps.H. BRIG analysis of sequences isolated from the lung indicated that the genes in the variable regions included cupC1 at locus 1050kbps, fliC at locus 1150kbps, pclass at locus 2400kbps, amiC at locus 3750kbps, PA3989 at locus 4450kbps and PA4108 at locus 4600kbps.I. BRIG analysis of sequences from the sputum indicated that the genes located in variable regions included p class at locus 2450kbps, PA2771 at locus 3150kbps, amiC at locus 3800kbps, PA4108 at locus 4600kbps, bfiS and bfiR at 4700kbps and PA4354 at 4900kbps.J. BRIG analysis for sequence from the trachea revealed that the genes in the variable regions included fliC at locus 1150kbps, p class at locus 2400kbs,
PA2771 at locus 3150kbps, amiC at locus 3750kbps, PA3989 at locus 4450kbps, bfiS and bfiR at locus 4700kbps, PA4625 at locus 5200kbps and algC at locus 6000kbps. K. BRIG analysis of sequences from urine samples indicated that the genes in variable regions included p class at 2400kbp and amiC at locus 3750kbps.L. BRIG analysis for sequences from wound infections indicated that the genes in the variable regions included p class at locus 2400kbps, PA2752 at locus 2850kbps and amiC at locus 3750kbps.M. BRIG analysis of sequences isolated from abscess indicated that genes in the variable regions included cupC1 at locus 1050kbps, fliC at locus 1150kbps, p class at locus 2400kbps, amiC at locus 3750kbps and algC at locus 6000kbps.N. BRIG analysis of sequences from blood isolates indicated that the genes present in the variable regions included p class at locus 2400kbps, PA2824 at locus 3150 kbps, amiC at locus 3750kbps, PA3989 at locus 4450kbps, PA4108 at locus 4600kbps, bfiS and bfiR at locus 4700kbps.