

# SCIENTIFIC REPORTS



OPEN

## Comparative genomics of clinical strains of *Pseudomonas aeruginosa* strains isolated from different geographic sites

Dinesh Subedi<sup>1</sup>, Ajay Kumar Vijay<sup>1</sup>, Gurjeet Singh Kohli<sup>2</sup>, Scott A. Rice<sup>2,3,4</sup> & Mark Willcox<sup>1</sup>

The large and complex genome of *Pseudomonas aeruginosa*, which consists of significant portions (up to 20%) of transferable genetic elements contributes to the rapid development of antibiotic resistance. The whole genome sequences of 22 strains isolated from eye and cystic fibrosis patients in Australia and India between 1992 and 2007 were used to compare genomic divergence and phylogenetic relationships as well as genes for antibiotic resistance and virulence factors. Analysis of the pangenome indicated a large variation in the size of accessory genome amongst 22 strains and the size of the accessory genome correlated with number of genomic islands, insertion sequences and prophages. The strains were diverse in terms of sequence type and dissimilar to that of global epidemic *P. aeruginosa* clones. Of the eye isolates, 62% clustered together within a single lineage. Indian eye isolates possessed genes associated with resistance to aminoglycoside, beta-lactams, sulphonamide, quaternary ammonium compounds, tetracycline, trimethoprim and chloramphenicols. These genes were, however, absent in Australian isolates regardless of source. Overall, our results provide valuable information for understanding the genomic diversity of *P. aeruginosa* isolated from two different infection types and countries.

The diverse and dynamic genetic composition of *Pseudomonas aeruginosa* enables this Gram-negative bacterium to colonise various environments, including humans where it can cause opportunistic infections<sup>1,2</sup>. *P. aeruginosa* is particularly associated with infections that are caused due to impaired anatomical structures or a weakened immune system. Such infections include microbial keratitis (MK), ventilator-associated pneumonia, wound infections, and respiratory infections in patients suffering from cystic fibrosis (CF)<sup>3–5</sup>. Several reports have shown that the prevalence of such infections by multidrug-resistant (MDR) strains is increasing rapidly worldwide<sup>6–9</sup>, which makes this bacterium difficult to treat and hence there is a high risk of mortality associated with infection by *P. aeruginosa*<sup>10</sup>. This pathogen has an exceptional capacity to develop resistance to antibiotics by the selection for genomic mutations and by exchange of transferable resistance determinants<sup>11</sup>. Knowledge of the genomic diversity of *P. aeruginosa* will help to understand differences in pathogenesis between strains and the mechanism of antibiotic resistance, which is important for controlling infections.

The genome size of *P. aeruginosa* varies greatly, ranging from 5.5 to 7 Mbp<sup>12,13</sup>. Such variation arises due to the presence of a large accessory genome. Accessory genomes are strain specific blocks of DNA and can occupy up to 20% of the whole genome<sup>14</sup>. They are composed of horizontally transferable elements which include prophages, transposons, insertion sequences (IS), genomic islands (GI) and plasmids<sup>15</sup>. Accessory genomes are important for carrying virulence and acquired antibiotic resistance genes. The lateral transfer of those genes between strains contributes to the development of MDR virulent strains<sup>16</sup>. Furthermore, mutational changes of chromosomal genes can also contribute to virulence and antibiotic resistance<sup>16,17</sup>. Therefore, unraveling the genetic content of *P. aeruginosa* helps to understand the gene modifications that are associated with more pathogenic and more resistant strains. Several studies have reported a comparison between genomes of *P. aeruginosa* in different infections at various points of time during infections<sup>16,18–21</sup>. However, most of those studies have centered around CF isolates and there is very limited comparative genomics of ocular isolates of *P. aeruginosa*.

<sup>1</sup>School of Optometry and Vision Science, University of New South Wales, Sydney, Australia. <sup>2</sup>Singapore Centre for Environmental Life Sciences Engineering, Singapore, Singapore. <sup>3</sup>The School of Biological Sciences, Nanyang Technological University, Singapore, Singapore. <sup>4</sup>The iThree Institute, The University of Technology Sydney, Sydney, NSW, Australia. Correspondence and requests for materials should be addressed to D.S. (email: [d.subedi@unsw.edu.au](mailto:d.subedi@unsw.edu.au))

	Strains	Sequence type <sup>†</sup>	No. of contigs	Length (bp)	GC (%)	CDS	tRNA	Accessory genes <sup>**</sup>
Eye/India	PA31	308	137	7100578	66.02	6619	69	1709
	PA32	308	155	7101589	66.01	6611	69	1701
	PA33	308	166	7092617	66.02	6609	69	1699
	PA34	1284	130	6885314	65.95	6326	66	1416
	PA35	308	156	7094960	66.02	6611	69	1701
	PA37	308	241	7154765	65.94	6645	69	1735
	PA82	1027	64	6387501	66.51	5810	65	900
								Average number of accessory genes = 1552
Eye/Australia	PA17	New	60	6360710	66.45	5825	72	915
	PA40	New	109	6284606	66.44	5700	69	790
	PA149	New	59	6314825	66.46	5745	68	835
	PA157	386	56	6249622	66.53	5708	68	798
	PA171	471	60	6339342	66.49	5812	69	902
	PA175	309	62	6757641	66.2	6181	68	1271
								Average number of accessory genes = 919
CF/Australia	PA55	549	77	6235554	66.57	5668	67	758
	PA57	New	73	6333117	66.48	5792	68	882
	PA59	New <sup>†</sup>	78	6289887	66.55	5767	68	857
	PA64	775	87	6264428	66.55	5713	65	803
	PA66	New <sup>†</sup>	93	6337310	66.51	5828	68	918
	PA86	New <sup>†</sup>	76	6170893	66.46	5685	68	775
	PA92	775	81	6144573	66.59	5584	65	674
	PA100	483	83	6310616	66.5	5732	66	822
	PA102	1717	62	6245474	66.55	5710	69	800
								Average number of accessory genes = 810
	PAO1*	549	1	6264404	66.6	5671	73	761

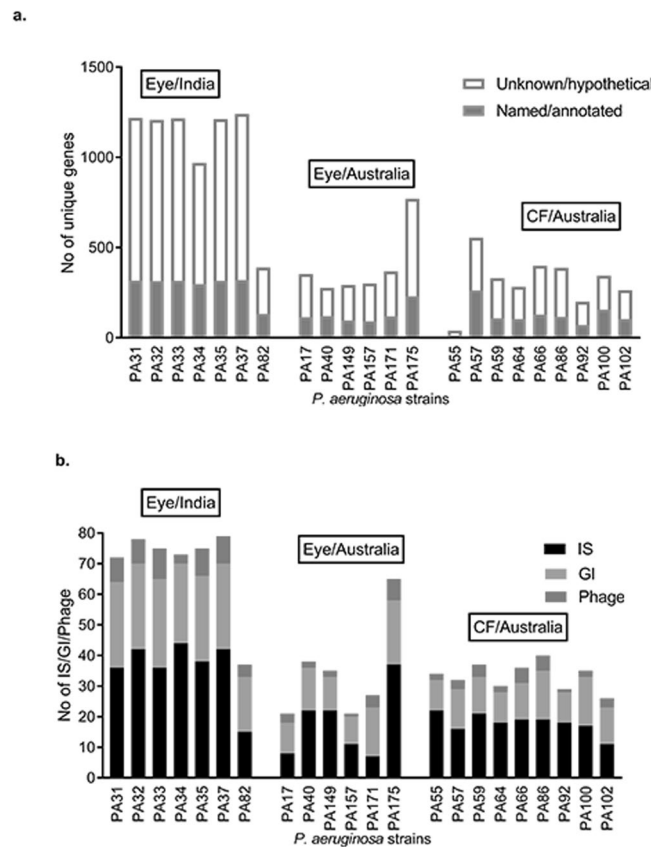
**Table 1.** General features of the genomes of *P. aeruginosa* strains. <sup>†</sup>Sequence types were determined by the multi locus sequence typing database. The sequence types not listed in the MLST database have been deemed as new. <sup>\*\*</sup>Accessory genes were determined by subtracting number of core genes (4910) from total number of CDS. <sup>†</sup>Same MLST allelic profile. \*Reference strain.

This study aims to compare the genomic diversity between *P. aeruginosa* strains from MK and CF isolated in Australia and India. There are previous reports of genomic characterisation of Indian ocular isolates of *P. aeruginosa*<sup>22–24</sup>. A genotypic study of eye isolates of *P. aeruginosa* has shown that keratitis isolates from the UK are highly related<sup>25</sup>. However, information on genomic comparison amongst contemporary isolates of *P. aeruginosa* from eye infections in different geographical locations is still missing. This study focussed on 13 MK strains, which were isolated in India and Australia and nine strains from CF cases which were isolated in Australia. The whole genomes of all 22 strains were sequenced and a comparative genomic analysis was conducted to identify genomic divergence, evolutionary relationships, antibiotic resistance properties and virulence factors.

## Results and Discussion

**General features of genomes.** A *de novo* assembly of the genomes of 22 *P. aeruginosa* strains generated a number of contigs from 56 in PA175 to 241 in PA37 (Median = 79). Like other published complete genomes of *P. aeruginosa*<sup>1,19,26,27</sup>, a mean C + G content of 66.4% and size of 6.1 to 7.1 Mbp was observed in the draft genomes. The genomic size varied widely between strains showing up to 900 kbp more DNA than PAO1, which was taken as the reference strain in this study. Similarly, the number of coding sequences (CDS), which were determined based on Prokka annotation pipeline, ranged from 5584 (in PA92) to 6645 (in PA37). Amongst 82 complete genomes of *P. aeruginosa* listed in the Pseudomonas genome database (PGDB)<sup>28</sup> (accessed on 12/03/2018), PA92 has the lowest and PA37 has the second highest number of CDS. Wide variations in the tRNA copy numbers (65–73) per strain observed here is probably due to use of incomplete draft genome. In addition, different number of tRNAs in the same genome was observed when annotated using different pipelines. Table 1 shows the general features of the genomes.

A total of 9786 orthologs were detected in all 22 strains and the reference strain PAO1. As the pan-genome represents the cumulative genetic information within a set of bacterial genomes, its size increases with the number and diversity of strains used for evaluation. A study that included 17 *P. aeruginosa* reference strains from diverse sources has found 9344 orthologs in the pangenome<sup>29</sup>, which is comparable to the results observed here. The higher number of genes in the pan-genome in our study may be the result of the diverse nature of the studied

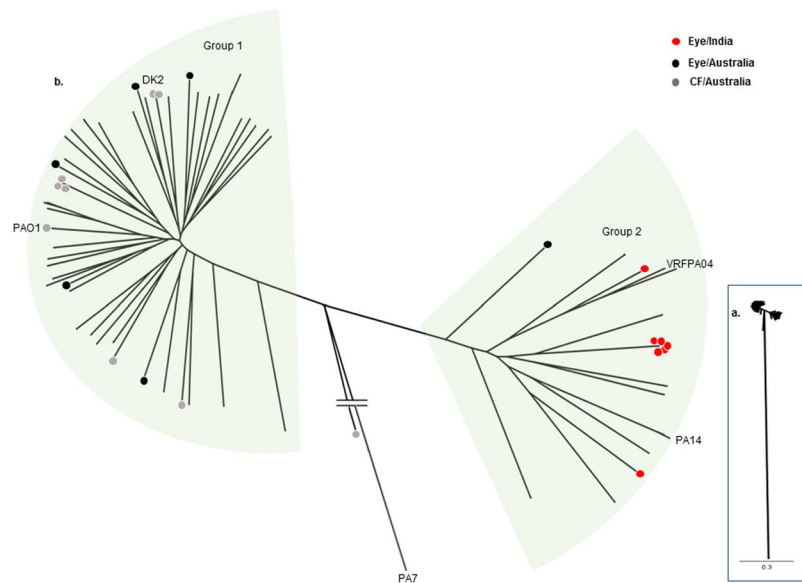


**Figure 1.** Composition of accessory genomes. (a) Distribution of unique genes. (b) Distribution of predicted no of insertion sequences (IS), genomic islands (GI) and phages.

strains. Out of the 9786 pan genes, 4910 genes were common in at least 99% of strains and this represents the core genome for the strains in the current study. Prior studies have reported core genomes of 5316<sup>30</sup>, 5233<sup>29</sup>, 5021<sup>31</sup>, and 4934<sup>14</sup> in different *P. aeruginosa* strains. Although the other studies used smaller sets (5 to 17) of genomes, the results are broadly comparable. Many factors may be responsible for the smaller core genomes in the current study including a larger population of genomes used for alignment, use of incomplete draft genomes, the diverse nature of the study populations (ocular and lung; Australian and Indian) and a strict definition of the core genome ( $\geq 99\%$  similarity in each strain). For example, pan-genome analysis of the same set of genomes of the current study but excluding PA57 and using  $\geq 95\%$  similarity resulted in 5287 core genes.

In addition to the large core genome, *P. aeruginosa* has accessory genomes that are not common in all strains<sup>15</sup>. The accessory genome can comprise of up to 20% of the total genome, and the majority of genes in this accessory genome are acquired horizontally. These genes include phages, transposons, IS and GI<sup>14</sup>. In the current study, the accessory genes were identified by subtracting the core genes (4910) from the total number of CDS. The frequency of accessory genes was 12% to 26%, which is more than the previously reported size of accessory genome<sup>29,32</sup>. However, the use of draft genomes may overestimate the number of accessory genes because of the presence of genomic repeats or transposable elements that may interrupt assembly and give an apparently larger genome than this actually present<sup>33</sup>. Accessory genomes may carry genes that help strains to persist in environments that may be unsuitable for others<sup>30</sup>. Like many other bacteria, the accessory genomes of *P. aeruginosa* encompasses genes related to virulence and antibiotic resistance<sup>34,35</sup>. The presence of a higher number of accessory genes in the set of ocular isolates indicates that eye strains may have acquired many genes to make this opportunistic species suitable to grow in the ocular environment. Furthermore, we examined the number of unique genes amongst accessory genes and found that the functions of the majority of the unique genes are unknown (Fig. 1a).

The genomes were examined for the presence of insertion sequences (IS), genomic islands (GI) and prophages, which are the main elements of an accessory genome<sup>15</sup>. Contigs of draft genomes were reordered with reference to PAO1 and the ordered contigs were joined together and made into a single FASTA file before examining databases. The results show that the average predicted number of GI was 26 (range 29–18) in Indian eye isolates, which was greater than that of Australian isolates (average 13) irrespective of source. Similarly, the average predicted number of IS and phages were higher in Indian eye isolates (Fig. 1b). Twenty (PA157) to 75 (PA33) total accessory elements were observed in all draft genomes. In contrast, a study has noted 38 to 53 accessory elements that are integrated into 89 potential genomic loci (region of genomic plasticity) in the complete genome of several *P. aeruginosa* strains<sup>14</sup>. Complete genomes are required to ascertain the actual number of genes in accessory genome. Nevertheless, the predicted number of IS, GI and phages was well correlated with the size of the accessory genome indicating that they contribute to the genomic diversity as highlighted in other studies<sup>15,35,36</sup>.



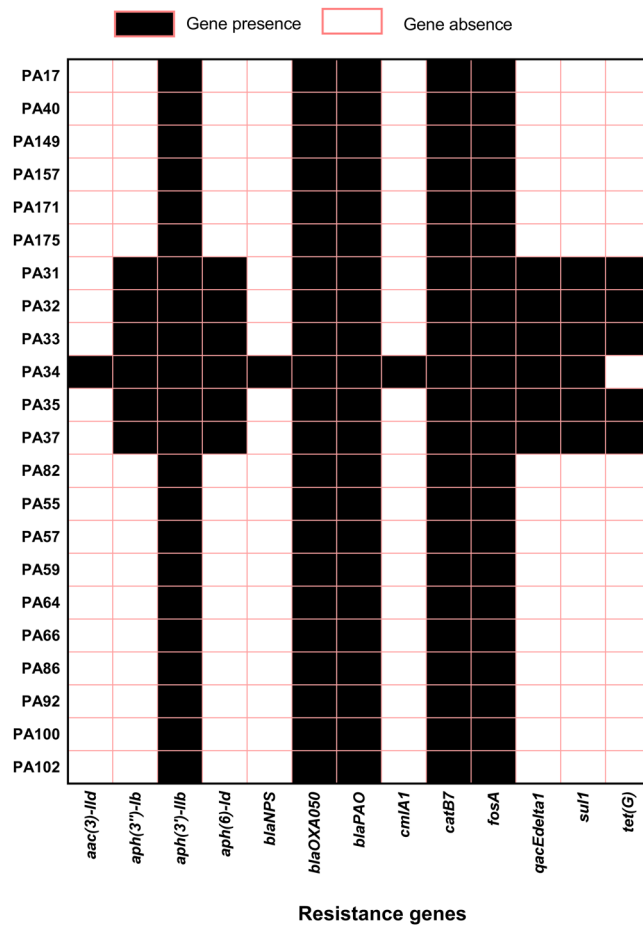
**Figure 2.** Phylogenetic analysis of *Pseudomonas aeruginosa* isolates. Maximum likelihood phylogenetic tree built with core genome SNPs based on mapping to the PAO1 excluding SNPs identified in regions that had arisen by recombination. **(a)** The original tree where the scale bars represent the number of substitutions per site. **(b)** Magnified tree showing branches and groups. Strains used in this study are indicated by distinct colour circles. Relative positions of few reference strains are shown, which are *P. aeruginosa* VRFP04, *P. aeruginosa* UCBPP-PA14, *P. aeruginosa* PAO1, and *P. aeruginosa* DK2.

From the pangenome and MLST analysis (below), five Indian strains, isolated from different patients with different histories, were found to be clonal and showed at least 99.98% sequence similarity with each other in MUMmer3<sup>37</sup> whole genome alignment. To avoid the overestimation of the accessory genome due to the dominance of a single clone, we obtained the nucleotide sequence of five additional Indian eye isolates from public databases<sup>22–24</sup> and reran the pan genome analysis. The relative size of the accessory genome to PAO1 was examined (Supplementary Fig. S1). The results tend to show that the eye isolates have larger accessory genome than CF isolates. However, due to limited number of clonally diverse strains of Indian origin, further research on larger datasets is required to confirm this.

Based upon MLST analysis, 16 distinct sequence types (STs) were found, with seven of these constituting new types. The ST was assigned to each strain according to the matched number in the MLST database<sup>38</sup>. Any strain that did not match with the existing database was deemed to have a new ST. Five Indian ocular isolates (out of seven) belonged to ST 308, two Australian CF isolates corresponded to ST 775 and three Australian CF strains had identical allelic profiles but did not match with any existing ST in the MLST database (all MLST profiles are shown in Supplementary Table S4). The remaining 13 STs were unique, with only a single representative (Table 1). Our results show that these strains belong to a diverse range of STs and are not similar to previously described clinical epidemic isolates<sup>39,40</sup>. Five strains with ST 308, collected from keratitis patients from the same centre in India, indicate the strains were potentially acquired from the same source where these strains may persist. The most common genotype observed in this study, ST 308, was also reported in MDR hospital strains in France<sup>39</sup>. Although the MLST database does not contain all *P. aeruginosa* strains, our observations show the diverse nature of the strains, which were not related to so-called world epidemic STs (ST 235, ST 111, ST 175, ST 395)<sup>39</sup>. This result also contradicts the previous finding that some keratitis isolates were clonally related with ST 235 CF strains<sup>41</sup>.

**Phylogenetics.** A total of 82 complete genomes of *P. aeruginosa* including PAO1 were downloaded from the NCBI database and used to compare the phylogenetic diversity of 22 strains from the current study. These 82 strains were listed in PGDB<sup>28</sup> as a complete genome and could represent a global *P. aeruginosa* collection. Core genome alignment was generated using Parsnp of the Harvest Suite with PAO1 as the reference. The alignment was then used to construct a tree following previously described methods, with *P. aeruginosa* PA7, a taxonomic outliner<sup>27</sup>, as an outgroup. A multi-sample variant call file was generated from the core genome alignment and SNPs present in all strains were examined (Supplementary Tables S1 and S2). In total 284,252 SNP sites were identified amongst 104 isolates.

All strains, except PA57, were clustered into two groups (Fig. 2). This is in agreement with several studies which have also shown that *P. aeruginosa* strains from various sources tend to cluster into two major groups<sup>42–44</sup>, with group 1 being larger, and which contains the most widely studied strains PAO1<sup>1</sup> and some notable CF strains such as DK2 and LESB58<sup>45,46</sup>. Group 2 tends to be smaller and includes the well known virulent strain PA14<sup>13</sup> and an Indian ocular isolate VRFP04, a virulent MDR strain<sup>24</sup>. All seven Indian and one Australian eye isolates were clustered into three sub-lineages within the group 2. A typing-based population structure analysis has also



**Figure 3.** The presence and absence of resistance genes as detected by Resfinder database. Associated resistance: Beta lactams: - *bla*<sub>OXA-50</sub>, *bla*<sub>PAO</sub>, *bla*<sub>NPS-1</sub> Aminoglycosides: - *aph(3'')-Ib*, *aph(6)-Id*, *aph(3')-Ib*, *aac(3)-IId* Fosfomycin: *fosA* Sulphonamide: *sul1* Chloramphenicol: *cmlA1*, *catB7* Tetracycline: *tet(G)*. Quaternary ammonium compounds: *qacEdelta1*.

unveiled that keratitis *P. aeruginosa* strains are closely related<sup>25</sup>. Furthermore, this supports the finding of the previous study that human *P. aeruginosa* are less diverse than isolates from the environment<sup>47</sup>. Similarly, all the CF strains and five Australian eye strains were of group 1 (See Supplementary Table S1 for phylogeny group classification of each strains and associated core genome SNPs). Amongst the CF isolates, continuous mutations have been shown to be an evolutionary process that may make a strain more pathogenic so that they rapidly transfer between humans<sup>16,21</sup>. However, previous studies have not focussed on ocular isolates. Our analysis showed that more than 60% of eye isolates clustered together in a single group, which is in agreement with previous findings that 71% of MK isolates of *P. aeruginosa* from the UK clustered together in the same group<sup>48</sup>. Further studies should focus on the evolutionary changes in ocular isolates of *P. aeruginosa* over a prescribed period of time. A CF strain PA57 was in a separate cluster and did not show similarity with other strains. This strain could be another taxonomic outlier of the *P. aeruginosa* (group 3)<sup>44</sup>.

**Antibiotic resistance gene profiles.** Horizontally acquired resistance genes were examined using the assembled contigs in the ResFinder database. Altogether, 13 distinct types of acquired resistance genes were detected in this study (Fig. 3). In common with other *P. aeruginosa* strains<sup>24</sup>, two beta-lactams (*bla*<sub>OXA-50</sub>, *bla*<sub>PAO</sub>) and one each for aminoglycoside (*aph(3'')-Ib*) fosfomycin (*fosA*) and chloramphenicol (*catB7*) resistance genes were present in all studied strains. Furthermore, six out of 22 strains had acquired additional resistance genes. Interestingly, all six strains were Indian eye isolates and possessed two aminoglycoside resistance genes (*aph(3'')-Ib* and *aph(6)-Id*), one sulphonamide resistance gene (*sul1*) and one quaternary ammonium compound resistance gene (*qacEdelta1*). The tetracycline efflux protein transporter gene *tet(G)* was detected in five Indian eye isolates, all of them are ST 308. An Indian eye strain PA34 possessed four unique resistance genes; *blaNPS-1*, *aac(3)-Ib*, *dfrA15* and *cmlA1* that can confer resistance to beta-lactams, aminoglycosides, trimethoprim and chloramphenicol, respectively. As horizontally acquired resistance genes may be associated with integrons, we analysed all of the draft sequences for the presence of integrons using Integron Finder version 1.5.1<sup>49</sup>. Although *sul1* and *qacEdelta1* are indicative of class I integrons, only strain PA34 possessed a class I integron, in agreement with a recent publication<sup>50</sup>. The acquired resistance genes detected were comparable to previous observations for an Indian eye isolate of *P. aeruginosa*<sup>24</sup>. As all Indian isolates of the current study were from the same centre

Strains		Antibiotics									
		Gentamicin	Ciprofloxacin	Levofloxacin	Moxifloxacin	Ceftazimidin	Cefepime	Imipenem	Ticarcillin	Aztronam	Polymyxin B
Eye/India	PA31	R	R	R	R	I	I	I	I	I	I
	PA32	R	R	R	R	I	S	I	I	S	I
	PA33	R	R	R	R	I	I	I	I	S	S
	PA34	R	I	S	R	S	I	R	I	I	S
	PA35	R	R	R	R	I	I	I	I	S	S
	PA37	R	R	R	R	I	S	I	I	S	S
	PA82	I	R	I	S	R	R	S	I	S	R
Eye/Australia	PA17	S	I	S	R	S	S	S	I	I	I
	PA40	S	R	S	S	S	S	I	I	S	S
	PA149	S	S	S	S	I	S	S	S	S	S
	PA157	S	S	S	S	I	S	S	I	S	S
	PA171	S	S	S	S	I	S	I	I	S	S
	PA175	S	I	S	S	I	S	I	I	S	S
CF/Australia	PA55	R	S	S	S	S	S	I	I	S	S
	PA57	R	S	S	S	S	S	S	I	S	S
	PA59	R	S	S	S	S	S	S	I	S	S
	PA64	I	R	R	R	S	R	S	I	S	S
	PA66	R	I	I	R	S	R	S	I	S	S
	PA86	S	S	S	S	I	S	S	I	S	S
	PA92	I	I	I	R	I	R	S	S	I	S
	PA100	I	I	S	R	I	S	S	S	S	S
	PA102	R	I	I	R	I	S	S	S	S	S

**Table 2.** Antibiotic susceptibility profile of *P. aeruginosa* strains.

in India, it is possible that there was antibiotic selection pressure that led to the selection for strains that had acquired such resistance genes from the environment. The absence of such genes in Australian isolates indicates that the antibiotic selection pressure may be different between Australia and India or that the genes associated with resistance are not readily accessible to *P. aeruginosa* in their local Australian environment. Furthermore, isolates from India were more likely to carry more resistance genes than Australian isolates, potentially reflecting the relatively unregulated use of antibiotics in India compared to Australia<sup>51</sup>. Antibiotic susceptibility tests also shows that Indian eye isolates were resistance to gentamicin and at least one fluoroquinolone. Resistance to aminoglycoside and fluoroquinolone is however, low in Australian isolates (Table 2).

On the basis of searches in the literature and online databases (Comprehensive Antibiotic Resistance Database (CARD), <https://card.mcmaster.ca/home> and the Pseudomonas genome database, <http://www.pseudomonas.com>), a set of 73 genes, which were related to antibiotic and disinfectant resistance in *P. aeruginosa*, were selected to examine variations in these genes between strains. Only high-quality, non-synonymous SNPs and indels were used for interpretation (Table 3). No insertions or deletions were detected in any of the strains. In terms of the number of SNPs and strains types, all Indian eye isolates and one Australian eye isolate (PA175) had relatively more variations (total SNPs >125) in the set of resistance genes than other strains. However, the CF strain PA57 had an exceptionally high number of SNPs in its resistome. Another CF strain, PA55, did not show any variations in its resistome. In terms of the total SNPs in resistance genes, the least number of variations ( $\leq 5$  SNPs) were found in five efflux pump-related genes (*oprM* (5) *cycB*(1) *mexF* (4) *nalD*(5) and *nfxB* (2)), three target alteration genes (*gyrB* (5) *tufA*(2) *tufB* (0)) and one inactivation gene *fosA* (3); these are highly conserved genes in *P. aeruginosa*.

**Virulence genes.** Virulence factors associated with keratitis and cystic fibrosis were selected based on the literature and published sequences in the Virulence Factor Data Base (VFDB)<sup>52</sup> to examine the presence or absence of genes related to pathogenicity in the strains. A dataset of 147 virulence genes of PAO1 associated with adherence, protease production, the type IV secretion system, quorum sensing, alginate production/regulation and toxins were curated from VFDB and used in BLAST searches (Fig. 4). For the *exoU* gene, PA14 was taken as the reference because it is not present in PAO1. All instances where there was an absence of a gene were manually examined with orthologs from the most widely studied strains recommended by the PGDB<sup>28</sup>. Out of 147 genes, variation in virulence genes were found for 20 genes. This was most evident for a set of effector proteins (toxins) related to the type III secretion system (*exoS*, *exoT*, *exoU* and *exoY*)<sup>53,54</sup>. As in previous studies, *exoS* was predominantly found in CF strains (present in eight out of nine strains) and *exoU* was primarily found in eye strains (present in eight out of 13 eye isolates)<sup>25,55–58</sup>. Furthermore, as determined by previous studies, *exoU* and *exoS* were mutually exclusive<sup>59</sup>. However, neither *exoU* nor *exoS* was detected in the CF strain PA57. As the *exoU* gene is carried by genomic islands<sup>53,60</sup>, *exoU* possessing strains showed larger accessory genomes and cluster together in the same phylogenetic group. The *exoT* (100%) and *exoY* (86%) genes were the most prevalent secretory toxins in the strains and this result is in agreement with previous findings<sup>61</sup>. In a recent study, *exoY* (55%) and *exoT* (5%)

Gene locus	Gene name	Mechanism	<i>P. aeruginosa</i> strains/number of SNPs																							Total SNPs
			PA 31	PA 32	PA 33	PA 34	PA 35	PA 37	PA 82	PA 17	PA 40	PA 149	PA 157	PA 171	PA 175	PA 55	PA 57	PA 59	PA 64	PA 66	PA 86	PA 92	PA 100	PA 102		
PA0156	<i>triA</i>	Antibiotic efflux	4	5	5	1	6	5	4	1		1			3		1							1	37	
PA0157	<i>triB</i>					1				1	1		2	1			1	1	1	1	1	1	1	1	1	14
PA0158	<i>triC</i>			2	2	2	1	2	2	1			1			1									1	18
PA0424	<i>mexR</i>			2	1	1	2	1	1	1						1		2								12
PA0425	<i>mexA</i>											1		1		1				1	1	1		1		7
PA0426	<i>mexB</i>			1	1	2		1	4									2		1			1			13
PA0427	<i>oprM</i>												1	1					1		1	1				5
PA1236	<i>farB</i>			1	1	1	1	1	1	1	1			1						1			1			11
PA1282	<i>lrfA</i>			6	6	9	4	6	8	6	8	5	5	6	4	4		7	3	4	3	3	4	3	4	108
PA1316	<i>lrfA</i>			2	2	2	1	2	2	3	3	2	1	3	2	3		3	2	2	2	2	2	2	4	47
PA1435	<i>mexM</i>			4	4	4	5	4	4	3	4	6	5	6	6	5		8	5	5	5	5	5	4	5	102
PA1436	<i>mdtC</i>			2	2	2	1	2	2	2	3	2		3	2	2			2	2	2	2	2	2	4	41
PA1754	<i>cysB</i>																							1		1
PA2018	<i>mexY</i>			5	5	5	5	5	5	3	1	1	1	4	1	3		3	1	1	1	1	1	1	2	55
PA2019	<i>mexX</i>			4	4	4	4	4	4	5	3	4	5	3	3	3		6	3	4	3	3	4	3	4	80
PA2389	<i>macA</i>			2	1	1	1	1	1	1		1		1	1	1		1	1		1	1		1	1	18
PA2390	<i>macB</i>			1	1	1	3	1	1	2	1	3	2	3	1	1		3	1	2	1	1	2	1	3	35
PA2391	<i>opmQ</i>			6	5	6	4	6	6	4	1	3	2	1	2	4		5	4	5	4	4	5	3	1	81
PA2491	<i>mexS</i>			2	2	2	1	2	2	2	1	1	1	1	1	1		4	2	2	2	2	1	1	1	34
PA2493	<i>mexE</i>						1				1				1	3		2	1	2	1	1	1			14
PA2494	<i>mexF</i>						1			1								2								4
PA2495	<i>oprN</i>			1	1	1	2	1	1	1		1	1			1		2		1			1			15
PA2525	<i>adeC</i>						3						5	1		1		2	2			2		3	1	20
PA2526	<i>muxC</i>										1				1			1	1		1	1				6
PA2527	<i>muxB</i>						1					2					1									4
PA2837	<i>opmA</i>			3	3	3	4	3	3	5		1	1		3	4		6	2	1	2	2	1	1	2	50
PA3019	<i>taeA</i>			1	1	1	3	1	1	1	1	1	2	2	2	2		4	1	1	1	1	1	3	1	32
PA3137	<i>farB</i>			1	1	1	1	1	1		1	2		2	2	2		2	1	1	1	1	1	1	1	24
PA3521	<i>opmE</i>			3	3	3	2	3	3	2	3	5	6	4	2	3		11	3	5	3	3	5	3	2	77
PA3522	<i>mexQ</i>			4	4	4	4	4	4	6	4	5	2	2	3	3		3	4	6	4	4	5	2	3	80
PA3523	<i>mexP</i>			2	2	2	1	2	2	2	3	2		2	1	2		3		3			3	2	2	36
PA3574	<i>nalD</i>										1				1					3						5
PA3676	<i>mexK</i>			1	1	1	2	1	1	5	1	2		2	2	3		6	4	2	4	4	2	2	2	48
PA3677	<i>mexJ</i>			2	2	2	2	2	2	2			1			2		3	3	3	3	3	3	1		36
PA3678	<i>mexL</i>			1	1	1	1	1	1	1	1		1	1		1		2	2	1	2	2	1			21
PA3894	<i>adeC</i>						1			1			1	1	1			2	2	1	2	2	1	1		16
PA4205	<i>mexG</i>			1	1	1		1	1	1								3								9
PA4206	<i>mexH</i>			1	1	1	2	1	1							1		1								9
PA4207	<i>mexI</i>			1	1	1	1	1	1	1	1			1	1	1			1	1	1	1	1			16
PA4208	<i>opmD</i>			3	3	3	3	3	3	2	5	1	1			4		6						1		38
PA4374	<i>mexV</i>			2	2	2	4	2	2	3		2	1	1	2	2		4	1	2	1	1	2	1	3	40
PA4375	<i>mexW</i>			2	2	2	1	2	2	3	3	1	2	3	1	1		2	1	1	1	1	1			32
PA4595	<i>yjkk</i>						1				1		2		1	2		1	3	1	2	1	1		1	17
PA4597	<i>oprJ</i>										2	2	2	2				9	3	3	5	3	3	3		37
PA4598	<i>mexD</i>			2	2	2	2	2	2	2	3	2	2	1	3	2		9	2	1	2	2	1	2	3	49
PA4599	<i>mexC</i>			7	8	8	3	8	8	6	4	1	4		1	5		9	1		1	1		4	4	83
PA4600	<i>nfxB</i>												1					1								2
PA4974	<i>opmH</i>			2	1	1	1	5	5			1			2	1		5	1	5	1	1	3	4		39
PA4990	<i>emrE</i>			1	1	1		1	2				1							1			1			9
PA4997	<i>msbA</i>		2	3	3	2	2	3	4						4		3		1			1	1		29	
PA5158	<i>adeC</i>		3	3	3		3	3	3	2	2	1	2		3		5	3	3	3		3	2	1	48	
PA5160	<i>farB</i>		4	3	3	2	4	3	5	3	3	3	3	3	3		3	3	4	3		3	4	3	65	
PA5518	<i>rosB</i>		3	3	3	2	3	3	3					1	1		1	1	1	1		1	1	1	29	

Continued

Gene locus	Gene name	Mechanism	<i>P. aeruginosa</i> strains/number of SNPs																							
			PA 31	PA 32	PA 33	PA 34	PA 35	PA 37	PA 82	PA 17	PA 40	PA 149	PA 157	PA 171	PA 175	PA 55	PA 57	PA 59	PA 64	PA 66	PA 86	PA 92	PA 100	PA 102	Total SNPs	
PA0706	<i>catB7</i>	Antibiotic inactivation	4	4	4	5	4	4	3	3	2	2	3	4		1	2	4	2	2	4	3	3	65		
PA1129	<i>fosA</i>										1					2								3		
PA4109	<i>ampR</i>		2	2	2	3	2	2	1				2	2		3								21		
PA4110	<i>ampC</i>		5	5	5	3	5	5	6	1	2	2	2	2	5	12	1	1	1	1	1	1	2	2	69	
PA4119	<i>Aph(3')-IIb</i>		2	2	2	3	2	2	2	1	1	1	1	4		3		1				1	1	2	31	
PA5514	<i>OXA-50</i>		1	2		1	3	2	5	2	3		2	3	3		4	1	3	1		3		4	43	
PA0004	<i>gyrB</i>	Antibiotic target alternation																	1	3	1			5		
PA0903	<i>alaS</i>		1	1	1	1	1	1	1		1						1		1	1			2	13		
PA1972	<i>pmrC</i>		3	3	3	3	3	3	1	1	1	4	1	1	2		3	2	3	2	2	3	4	1	49	
PA3002	<i>mfd</i>		1	2	2	3	2	2	1	1	1		2		1		2	1	2	1	1	2			27	
PA3168	<i>gyrA</i>		1	1	1		1	1							2		2		1					1	11	
PA3946	<i>rosC</i>		6	8	8	3	7	6	1	1	1	1	2	2				1	2	1	1	2	2	1	56	
PA4265	<i>tufA</i>		1					1																	2	
PA4277	<i>tufB</i>																								0	
PA4560	<i>ileS</i>		2	2	2	2	2	2	2	1	4	2	1	1	2		4	2	4	2	2	4	3	1	47	
PA4964	<i>parC</i>		2	2	2		2	2			1				1		1								13	
PA4967	<i>parE</i>		1	1	1	1	1	1	2	1					1		1	1	1	1	1	2		1	18	
PA3553	<i>pmrF</i>																3	1		1	1		1		7	
PA3554	<i>arnA</i>		2	4	4	5	4	4	3	2	3		3		4		6		2			2	3	4	55	
PA0920	<i>mprF</i>		6	6	6	4	6	6	9		2	1	1	2	7		8	2	1	2	2	1	2	2	76	
			Total SNPs	137	140	144	124	146	150	136	82	89	79	83	77	125	0	217	88	109	88	81	101	87	86	

**Table 3.** Non-synonymous SNPs detected in the 73 genes related to antibiotic resistance in the 22 isolates studied using PAO1 as the reference genome.

were less prevalent than in the current study although the reason for these differences in distribution remains unclear<sup>62</sup>. One possible reason for this difference is that the study examined genes on the basis of PCR products, which may not be able to capture all different orthologs of genes.

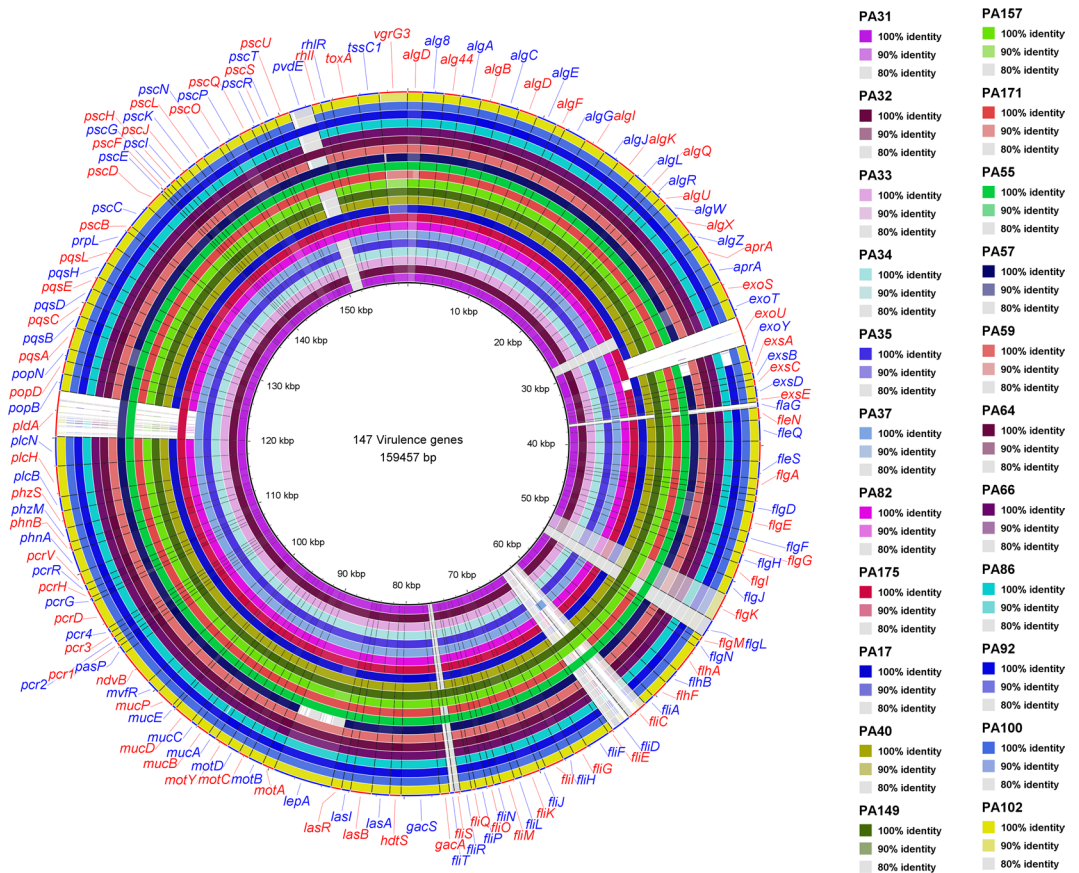
Flagellar genes help in the establishment of infections as they can be involved in adherence to surfaces and were also widely variable between strains<sup>63</sup>. Seven flagellar genes (*flgK*, *flgL*, *fliC*, *flaG*, *fliD*, *fliS*, and *fliT*) clustered between PA1086 and PA1096 in PAO1 were not matched with those of 19 strains that included both eye and CF isolates. However, these genes from 19 strains showed 90% to 99% similarity with genes between PA7\_4275 and PA7\_4291 of PA7, orthologs of the above seven flagellar genes. There was low sequence similarity (<50%) for the above flagellar genes between PAO1 and PA7. Studies involving CF isolates have shown that the activity of the *fliC* gene (that encodes flagellin) had been either downregulated<sup>64</sup> or was absent in some strains<sup>63</sup>. As flagella are immunogenic, the loss of flagella may be an important antiphagocytic mechanism in chronic infection isolates<sup>65</sup>. Although it has been shown that non-flagellated strains are defective in acute infections<sup>65</sup>, 85% of eye isolates in this study had altered flagellar genes that may affect flagellar function. Previous work has shown that although *fliC* contributes to invasion of *P. aeruginosa* in eye infections, a lack of *fliC* did not cause complete loss of invasion<sup>66</sup>. Further studies will need to clarify the functionality of those flagellar genes on studied strains and their role in ocular *P. aeruginosa* infections.

A phospholipase D gene (*pldA*), a part of the type VI secretion system of *P. aeruginosa* is believed to promote chronic infections<sup>67,68</sup>. However, *pldA* was absent from 13 isolates, seven of which were CF isolates and yet over 50% eye isolates had *pldA*. Reports on the role of *pldA* in eye infections have not been published and this should be an area of future study. Another notable variation was observed in *pvdE*, a precursor for pyoverdinin synthesis, which is essential for virulence of *P. aeruginosa*<sup>69,70</sup>. Eight strains, irrespective of their source of isolation, had a PAO1 homolog of *pvdE*. Similarly, DK2 and LES homologs of *pvdE* were equally distributed in 14 strains (Table 4) suggesting that these orthologs are evenly distributed in *P. aeruginosa* populations. PvdE can increase invasion of *P. aeruginosa* by inducing expression of the *exoS*<sup>71</sup>. Further studies will help understand role of *pvdE* variants in pathogenesis.

## Conclusions

This study compared the genomic variations between Australian and Indian *P. aeruginosa* isolates from ocular infections. *P. aeruginosa* isolates from various sources showed diversity in the size of accessory genome, antibiotic resistance genes and virulence factors. We found a slightly smaller core genome than has been reported previously. Although all 22 strains were distributed throughout the global phylogeny of *P. aeruginosa*, certain clusters were observed in the eye isolates where five Indian eye isolates were clustered into a single clonal lineage in the group which also contains a well-studied and virulent strain PA14. Larger accessory genomes were associated with eye isolates of this group. Furthermore, the strains of this group had more SNPs in their set of 73 resistome suggesting possible positive antibiotic selection pressure. Variation in virulence factors, except for *exoU*, was not





**Figure 4.** A circular representation of the genomes of studied isolates. The draft genomes of 22 strains were aligned against the 147 virulence genes curated from VFDB. Each genome is represented by a ring with different colours, which are shown in figure. Image was generated using BRIG (<http://brig.sourceforge.net>).

Strains	<i>pvdE</i> orthologs (locus tag)
PA31, PA32, PA33, PA34, PA35, PA37 and PA175	<i>P. aeruginosa</i> DK2 (DK2_13280)
PA82, PA17, PA171, PA175, PA55, PA57, PA64, and PA92	<i>P. aeruginosa</i> PAO1 (PA2397)
PA40, PA149, PA59, PA66, PA86, PA100 and PA102	<i>P. aeruginosa</i> LESB58 (PALES_28991)

**Table 4.** Distribution of *pvdE* orthologs among strains.

correlated with phylogeny. This study relied on draft genomes and may not be able to predict actual genomic diversity because the analysis could not ascertain the presence of the plasmids in any of the isolates. Further studies will focus on improvement of the assembly of these genomes. Overall, these findings have extended our understanding of the genomic diversity of *P. aeruginosa* in two different infections and information can be used to elucidate various mechanism that would help fight against virulent and drug resistant strains.

## Methods

**Bacterial strains and antibiotic susceptibility tests.** Twenty two clinical isolates of *P. aeruginosa* from corneas of microbial keratitis and from the lungs of CF patients were selected for this study. Seven ocular isolates were obtained from a tertiary eye care centre in India (L.V. Prasad Eye Institute, Hyderabad, India), six ocular and nine CF isolates were acquired from various sources in Australia. All strains were collected from institutional repositories between 1992 and 2007 without identifiable patient data and all experiments followed the institutional guidelines, which were in place at the time (Table 5). Genome sequence data of an additional 82 *P. aeruginosa* strains, based on availability of complete genome sequence in Pseudomonas genome database (PGDB) version 17.2<sup>28</sup> including *P. aeruginosa* PAO1 (reference strain) were collected from public databases and used in this study to compare results and to build phylogenetic trees (all the reference strains used in this study are listed in Supplementary Table S1). The minimum inhibitory concentrations (MICs) of ceftazidime, cefepime, aztreonam, ticarcillin, imipenem, gentamicin, levofloxacin, ciprofloxacin, moxifloxacin and polymyxin were determined by broth microdilution according to CLSI guidelines and published standard breakpoints<sup>72–74</sup>.

Strains	Collection date <sup>a</sup>	Geographical location	Associated infections
PA31	02/10/1997	LVPEI, Hyderabad, India	Microbial Keratitis
PA32	08/10/1997	LVPEI, Hyderabad, India	Microbial Keratitis
PA33	29/08/1997	LVPEI, Hyderabad, India	Microbial Keratitis
PA34	28/08/1997	LVPEI, Hyderabad, India	Microbial Keratitis
PA35	09/08/1997	LVPEI, Hyderabad, India	Microbial Keratitis
PA37	11/07/1997	LVPEI, Hyderabad, India	Microbial Keratitis
PA82	11/05/2004	LVPEI, Hyderabad, India	Microbial Keratitis
PA17	15/09/1992	Flinders, Adelaide, Australia	Microbial Keratitis
PA40	02/02/1999	SEH, Sydney, Australia	Microbial Keratitis
PA149	04/03/2004	Flinders, Adelaide, Australia	Microbial Keratitis
PA157	29/04/2006	PAH, Brisbane, Australia	Microbial Keratitis
PA171	16/03/2006	PAH, Brisbane, Australia	Microbial Keratitis
PA175	07/10/2006	PAH, Brisbane, Australia	Microbial Keratitis
PA55	2003	RPAH, Sydney, Australia	Cystic Fibrosis
PA57	2003	RPAH, Sydney, Australia	Cystic Fibrosis
PA59	2003	RPAH, Sydney, Australia	Cystic Fibrosis
PA64	2003	RPAH, Sydney, Australia	Cystic Fibrosis
PA66	2003	RPAH, Sydney, Australia	Cystic Fibrosis
PA86	2004	RPAH, Sydney, Australia	Cystic Fibrosis
PA92	2004	RPAH, Sydney, Australia	Cystic Fibrosis
PA100	2004	RPAH, Sydney, Australia	Cystic Fibrosis
PA102	2004	RPAH, Sydney, Australia	Cystic Fibrosis

**Table 5.** List of strains used in this study. LVPEI = LV Prasad Eye Institute; Flinders = Flinders University, SEH = Sydney Eye Hospital; PAH = Princes Alexandra Hospital; RPAH = Royal Prince Alfred Hospital CF Clinic, Sydney, Australia. <sup>a</sup>All cystic fibrosis isolates were obtained from Royal Prince Alfred Hospital CF Clinic, Sydney, Australia, between 2003 and 2004. Information on exact date of collection is missing in our record.

**Whole genome sequencing.** Genomic DNA was extracted from overnight cultures using the DNeasy<sup>®</sup> Blood and Tissue Kit (QIAGEN<sup>®</sup>, Germany) following the manufacturer's instructions. The paired-end library was prepared using Nextera XT DNA library preparation kit (Illumina<sup>®</sup>, San Diego, CA, USA). Libraries were then sequenced on Illumina<sup>®</sup> MiSeq bench top sequencer (Illumina), generating 300 bp paired-end reads. All of the libraries were multiplexed on one MiSeq run.

**Genome assembly and sequence analysis.** The MiSeq sequencing resulted an average of 760,773 reads (range 632,180 to 1,193,844) per isolate. FastQC version 0.11.7 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>) was used to assess the quality of raw reads, which were then quality trimmed to remove adaptor sequences using Trimmomatic version 0.36 and with the setting of minimum read length of 36 and minimum coverage of 15<sup>75</sup>. A *de novo* assembly was performed by SPAdes version 3.11.1<sup>76</sup>. with the default setting. The annotations of the assembled genomes were performed using Prokka version 1.7. using GenBank<sup>®</sup> compliance flag<sup>77</sup>. The genome of *P. aeruginosa* PAO1 (RefSeq accession number NC\_002516.2), which was used as the reference in this study, was re-annotated with Prokka to avoid annotation bias. Whenever necessary, the contigs of the draft genomes were reordered and/or aligned with the reference genome using MAUVE multiple-genome alignment software<sup>78,79</sup>. Artemis, a genome browser tool<sup>80</sup>, was used to concatenate the ordered contigs to get a single fragment of genomes which were used to examine insertion sequence using web tool ISSaga ([http://issaga.biotoul.fr/ISSaga2/issaga\\_index.php](http://issaga.biotoul.fr/ISSaga2/issaga_index.php)), genomic islands using IslandViewer 4<sup>81</sup> and prophages using PHASTER<sup>82</sup> Multi locus sequence type (MLST) was determined using pubMLST database<sup>38</sup> to find sequence type (ST) of each strain.

**Pan-genome and Phylogenetics.** The pangenome analysis was performed using Roary version 3.12.0<sup>83</sup> which uses the GFF3 files produced by Prokka. The program was run using the default settings, which uses BLASTp for all-against-all comparison with a percentage sequence identity of 95%. Core-genes were taken as the genes which were common in at least 99% of strains. The accessory genome was obtained as the genes present in the genome of each strain minus core genes. The Roary “gene\_presence\_absence.csv” file was further examined for unique genes using “union” and “difference” command. Parsnp version 1.2 in the Harvest Suite<sup>84</sup> was used to align the genomes of 104 *P. aeruginosa* strains (82 complete genomes from the PGDB and 22 draft genomes from this study), followed by the construction of a maximum likelihood tree based on core genome single nucleotide polymorphisms (SNPs), excluding SNPs identified in regions that had arisen by recombination.

**Variant calling.** The paired-end reads for each isolate were aligned against the genome of the *P. aeruginosa* PAO1 using Bowtie2 version 2.3.2<sup>85</sup> following “score-min” command to avoid alignments that score less than the default minimum score threshold and with “local” flag for better score. Genomic variants were compiled using “mpileup” in SAMtools, version 1.7<sup>86</sup>. A minimum quality score of 50 was set to list the SNPs and Indels. The

genomic variants were annotated using SnpEff version 4.3<sup>87</sup> with the default options to obtain the nucleotide changes and the predicted effects at the protein level.

**Antibiotic resistance and virulence genes.** Genomes were examined for the presence of acquired resistance genes using Resfinder 3.0 (Centre for Genomic Epidemiology, DTU, Denmark)<sup>88</sup>. Furthermore, a set of 73 genes related to antibiotic and disinfectant resistance in *P. aeruginosa* were selected from searches in the online databases Comprehensive Antibiotic Resistance Database (CARD) (<https://card.mcmaster.ca/home>)<sup>89</sup> and Pseudomonas genome database (<http://www.pseudomonas.com>)<sup>28</sup>. These 73 genes were manually examined for the presence of non-synonymous SNPs to predict genotypic changes in the resistome (see Supplementary Table S3).

A dataset of 146 virulence genes of PAO1 and one virulence gene (*exoU*) of *Pseudomonas aeruginosa* UCBPP-PA14 (NC\_008463.1) associated with adherence (flagella), protease production, type IV secretion system, quorum sensing, alginate production/regulation and toxins were curated from the Virulence Factor Data Base (VFDB)<sup>90</sup> and used in BLAST searches to match them with the genomes of the strains studied here. BLAST Ring Image Generator (BRIG)<sup>91</sup> was used to generate an image that shows presence or absence of virulence genes in multiple genomes. (List of virulence genes used is shown in Supplementary Table S4). The contigs were joined together before searching them in BRIG to avoid false matching due to fragmented genomes. The absence of a gene in this analysis was confirmed by manual BLASTn searching using orthologs from a widely-studied panel of *P. aeruginosa* suggested by PGDB. These include PA14, *P. aeruginosa* LESB58 (NC\_011770.1), *P. aeruginosa* PA7 (NC\_009656.1) and *P. aeruginosa* DK2 (CP003149.1).

**Nucleotide accession.** The nucleotide sequences are available in the GenBank under the Bioproject accession number PRJNA431326.

## References

1. Stover, C. K. *et al.* Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**, 959–964, <https://doi.org/10.1038/35023079> (2000).
2. Richards, M. J., Edwards, J. R., Culver, D. H. & Gaynes, R. P. Nosocomial infections in medical intensive care units in the United States. National Nosocomial Infections Surveillance System. *Crit Care Med* **27**, 887–892 (1999).
3. Stapleton, F. & Carnt, N. Contact lens-related microbial keratitis: how have epidemiology and genetics helped us with pathogenesis and prophylaxis. *Eye (Lond)* **26**, 185–193, <https://doi.org/10.1038/eye.2011.288> (2012).
4. Lyczak, J. B., Cannon, C. L. & Pier, G. B. Lung infections associated with cystic fibrosis. *Clin Microbiol Rev* **15**, 194–222 (2002).
5. Vaez, H. *et al.* Efflux pump regulatory genes mutations in multidrug resistance *Pseudomonas aeruginosa* isolated from wound infections in Isfahan hospitals. *Adv Biomed Res* **3**, 117, <https://doi.org/10.4103/2277-9175.133183> (2014).
6. Ding, C. *et al.* Prevalence of *Pseudomonas aeruginosa* and antimicrobial-resistant *Pseudomonas aeruginosa* in patients with pneumonia in mainland China: a systematic review and meta-analysis. *International Journal of Infectious Diseases* **49**, 119–128, <https://doi.org/10.1016/j.ijid.2016.06.014> (2016).
7. Gad, G. F., El-Domany, R. A., Zaki, S. & Ashour, H. M. Characterization of *Pseudomonas aeruginosa* isolated from clinical and environmental samples in Minia, Egypt: prevalence, antibiogram and resistance mechanisms. *J Antimicrob Chemother* **60**, 1010–1017, <https://doi.org/10.1093/jac/dkm348> (2007).
8. Kouda, S. *et al.* Increased prevalence and clonal dissemination of multidrug-resistant *Pseudomonas aeruginosa* with the blaIMP-1 gene cassette in Hiroshima. *J Antimicrob Chemother* **64**, 46–51, <https://doi.org/10.1093/jac/dkp142> (2009).
9. Mahmoud, A., Zahran, W., Hindawi, G., Labib, A. & Galal, R. Prevalence of multidrug-resistant *Pseudomonas aeruginosa* in patients with nosocomial infections at a University hospital in Egypt, with special reference to typing methods. *Journal of Virology & Microbiology*, 1–13, <https://doi.org/10.5171/2013.290047> (2013).
10. Pena, C. *et al.* Prospective multicenter study of the impact of carbapenem resistance on mortality in *Pseudomonas aeruginosa* bloodstream infections. *Antimicrob Agents Chemother* **56**, 1265–1272, <https://doi.org/10.1128/AAC.05991-11> (2012).
11. Livermore, D. M. Multiple mechanisms of antimicrobial resistance in *Pseudomonas aeruginosa*: our worst nightmare? *Clin Infect Dis* **34**, 634–640, <https://doi.org/10.1086/338782> (2002).
12. Schmidt, K. D., Tummler, B. & Romling, U. Comparative genome mapping of *Pseudomonas aeruginosa* PAO with *P. aeruginosa* C, which belongs to a major clone in cystic fibrosis patients and aquatic habitats. *J Bacteriol* **178**, 85–93 (1996).
13. Lee, D. G. *et al.* Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biol* **7**, R90, <https://doi.org/10.1186/gb-2006-7-10-r90> (2006).
14. Klockgether, J., Cramer, N., Wiehlmann, L., Davenport, C. F. & Tummler, B. *Pseudomonas aeruginosa* Genomic Structure and Diversity. *Front Microbiol* **2**, 150, <https://doi.org/10.3389/fmicb.2011.00150> (2011).
15. Kung, V. L., Ozer, E. A. & Hauser, A. R. The accessory genome of *Pseudomonas aeruginosa*. *Microbiol Mol Biol Rev* **74**, 621–641, <https://doi.org/10.1128/MMBR.00027-10> (2010).
16. Jeukens, J. *et al.* Comparative genomics of isolates of a *Pseudomonas aeruginosa* epidemic strain associated with chronic lung infections of cystic fibrosis patients. *PLoS One* **9**, e87611, <https://doi.org/10.1371/journal.pone.0087611> (2014).
17. Ashish, A., Shaw, M., Winstanley, C., Ledson, M. J. & Walshaw, M. J. Increasing resistance of the Liverpool Epidemic Strain (LES) of *Pseudomonas aeruginosa* (Psa) to antibiotics in cystic fibrosis (CF)—a cause for concern? *J Cyst Fibros* **11**, 173–179, <https://doi.org/10.1016/j.jcf.2011.11.004> (2012).
18. De Soya, A. *et al.* Developing an international *Pseudomonas aeruginosa* reference panel. *Microbiologyopen* **2**, 1010–1023, <https://doi.org/10.1002/mbo.3.141> (2013).
19. Jeukens, J., Kukavica-Ibrulj, I., Emond-Rheault, J. G., Freschi, L. & Levesque, R. C. Comparative genomics of a drug-resistant *Pseudomonas aeruginosa* panel and the challenges of antimicrobial resistance prediction from genomes. *FEMS Microbiol Lett* **364**, <https://doi.org/10.1093/femsle/fnx161> (2017).
20. Klockgether, J. *et al.* Intracolon diversity of the *Pseudomonas aeruginosa* cystic fibrosis airway isolates TBCF10839 and TBCF121838: distinct signatures of transcriptome, proteome, metabolome, adherence and pathogenicity despite an almost identical genome sequence. *Environ Microbiol* **15**, 191–210, <https://doi.org/10.1111/j.1462-2920.2012.02842.x> (2013).
21. Marvig, R. L., Johansen, H. K., Molin, S. & Jelsbak, L. Genome analysis of a transmissible lineage of *Pseudomonas aeruginosa* reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genet* **9**, e1003741, <https://doi.org/10.1371/journal.pgen.1003741> (2013).
22. Aggarwal, R. K., Dawar, C., Das, S. & Sharma, S. Draft genome sequences of two drug-resistant isolates of *Pseudomonas aeruginosa* obtained from keratitis patients in India. *Genome Announc* **3**, <https://doi.org/10.1128/genomeA.01404-14> (2015).

23. Murugan, N., Malathi, J., Umashankar, V. & Madhavan, H. N. Resistome and pathogenomics of multidrug resistant (MDR) *Pseudomonas aeruginosa* VRFPA03, VRFPA05 recovered from alkaline chemical keratitis and post-operative endophthalmitis patient. *Gene* **578**, 105–111, <https://doi.org/10.1016/j.gene.2015.12.022> (2016).
24. Murugan, N., Malathi, J., Umashankar, V. & Madhavan, H. N. Unraveling genomic and phenotypic nature of multidrug-resistant (MDR) *Pseudomonas aeruginosa* VRFPA04 isolated from keratitis patient. *Microbiol Res* **193**, 140–149, <https://doi.org/10.1016/j.micres.2016.10.002> (2016).
25. Stewart, R. M. *et al.* Genetic characterization indicates that a specific subpopulation of *Pseudomonas aeruginosa* is associated with keratitis infections. *J Clin Microbiol* **49**, 993–1003, <https://doi.org/10.1128/JCM.02036-10> (2011).
26. Miyoshi-Akiyama, T., Kuwahara, T., Tada, T., Kitao, T. & Kirikae, T. Complete genome sequence of highly multidrug-resistant *Pseudomonas aeruginosa* NCGM2.S1, a representative strain of a cluster endemic to Japan. *J Bacteriol* **193**, 7010, <https://doi.org/10.1128/JB.06312-11> (2011).
27. Roy, P. H. *et al.* Complete genome sequence of the multiresistant taxonomic outlier *Pseudomonas aeruginosa* PA7. *PLoS One* **5**, e8842, <https://doi.org/10.1371/journal.pone.0008842> (2010).
28. Winsor, G. L. *et al.* Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res* **44**, D646–653, <https://doi.org/10.1093/nar/gkv1227> (2016).
29. Valot, B. *et al.* What It Takes to Be a *Pseudomonas aeruginosa*? The Core Genome of the Opportunistic Pathogen Updated. *PLoS One* **10**, e0126468, <https://doi.org/10.1371/journal.pone.0126468> (2015).
30. Ozer, E. A., Allen, J. P. & Hauser, A. R. Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt. *BMC Genomics* **15**, 737, <https://doi.org/10.1186/1471-2164-15-737> (2014).
31. Mathee, K. *et al.* Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proc Natl Acad Sci USA* **105**, 3100–3105, <https://doi.org/10.1073/pnas.0711982105> (2008).
32. Pohl, S. *et al.* The extensive set of accessory *Pseudomonas aeruginosa* genomic components. *FEMS Microbiol Lett* **356**, 235–241, <https://doi.org/10.1111/1574-6968.12445> (2014).
33. Ricker, N., Qian, H. & Fulthorpe, R. R. The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. *Genomics* **100**, 167–175, <https://doi.org/10.1016/j.ygeno.2012.06.009> (2012).
34. Ho Sui, S. J., Fedynak, A., Hsiao, W. W. L., Langille, M. G. I. & Brinkman, F. S. L. The Association of Virulence Factors with Genomic Islands. *PLoS ONE* **4**, e8094, <https://doi.org/10.1371/journal.pone.0008094> (2009).
35. Shen, K. *et al.* Extensive genomic plasticity in *Pseudomonas aeruginosa* revealed by identification and distribution studies of novel genes among clinical isolates. *Infect Immun* **74**, 5272–5283, <https://doi.org/10.1128/iai.00546-06> (2006).
36. Roy Chowdhury, P. *et al.* Genomic islands 1 and 2 play key roles in the evolution of extensively drug-resistant ST235 isolates of *Pseudomonas aeruginosa*. *Open Biol* **6**, <https://doi.org/10.1098/rsob.150175> (2016).
37. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biology* **5**, <https://doi.org/10.1186/gb-2004-5-2-r12> (2004).
38. Jolley, K. A. & Maiden, M. C. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11**, 595, <https://doi.org/10.1186/1471-2105-11-595> (2010).
39. Cholley, P. *et al.* Most multidrug-resistant *Pseudomonas aeruginosa* isolates from hospitals in eastern France belong to a few clonal types. *J Clin Microbiol* **49**, 2578–2583, <https://doi.org/10.1128/JCM.00102-11> (2011).
40. Chen, S. H., Chen, R. Y., Xu, X. L. & Chen, H. T. Multilocus sequencing typing of *Pseudomonas aeruginosa* isolates and analysis of potential pathogenicity of typical genotype strains from occupational oxyhelium saturation divers. *Undersea Hyperb Med* **41**, 135–141 (2014).
41. Hall, A. J. *et al.* Intraclonal genetic diversity amongst cystic fibrosis and keratitis isolates of *Pseudomonas aeruginosa*. *J Med Microbiol* **62**, 208–216, <https://doi.org/10.1099/jmm.0.048272-0> (2013).
42. Stewart, L. *et al.* Draft genomes of 12 host-adapted and environmental isolates of *Pseudomonas aeruginosa* and their positions in the core genome phylogeny. *Pathog Dis* **71**, 20–25, <https://doi.org/10.1111/2049-632X.12107> (2014).
43. Kos, V. N. *et al.* The resistome of *Pseudomonas aeruginosa* in relationship to phenotypic susceptibility. *Antimicrob Agents Chemother* **59**, 427–436, <https://doi.org/10.1128/AAC.03954-14> (2015).
44. Freschi, L. *et al.* Clinical utilization of genomics data produced by the international *Pseudomonas aeruginosa* consortium. *Front Microbiol* **6**, 1036, <https://doi.org/10.3389/fmicb.2015.01036> (2015).
45. Wiehlmann, L., Cramer, N. & Tummler, B. Habitat-associated skew of clone abundance in the *Pseudomonas aeruginosa* population. *Environ Microbiol Rep* **7**, 955–960, <https://doi.org/10.1111/1758-2229.12340> (2015).
46. Rau, M. H., Marvig, R. L., Ehrlich, G. D., Molin, S. & Jelsbak, L. Deletion and acquisition of genomic content during early stage adaptation of *Pseudomonas aeruginosa* to a human host environment. *Environ Microbiol* **14**, 2200–2211, <https://doi.org/10.1111/j.1462-2920.2012.02795.x> (2012).
47. Winstanley, C. *et al.* Newly introduced genomic prophage islands are critical determinants of *in vivo* competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. *Genome Res* **19**, 12–23, <https://doi.org/10.1101/gr.086082.108> (2009).
48. Shankar, J. *et al.* Genotypic analysis of UK keratitis-associated *Pseudomonas aeruginosa* suggests adaptation to environmental water as a key component in the development of eye infections. *FEMS Microbiol Lett* **334**, 79–86, <https://doi.org/10.1111/j.1574-6968.2012.02621.x> (2012).
49. Curry, J., Jove, T., Touchon, M., Neron, B. & Rocha, E. P. Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Res* **44**, 4539–4550, <https://doi.org/10.1093/nar/gkw319> (2016).
50. Subedi, D., Vijay, A. K., Kohli, G. S., Rice, S. A. & Willcox, M. Nucleotide sequence analysis of NPS-1 beta-lactamase and a novel integron (In1427)-carrying transposon in an MDR *Pseudomonas aeruginosa* keratitis strain. *J Antimicrob Chemother*, <https://doi.org/10.1093/jac/dky073> (2018).
51. Laxminarayan, R. *et al.* Access to effective antimicrobials: a worldwide challenge. *The Lancet* **387**, 168–175, [https://doi.org/10.1016/S0140-6736\(15\)00474-2](https://doi.org/10.1016/S0140-6736(15)00474-2) (2016).
52. Chen, L. *et al.* VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Research* **33**, D325–D328, <https://doi.org/10.1093/nar/gki008> (2005).
53. Sato, H. *et al.* The mechanism of action of the *Pseudomonas aeruginosa*-encoded type III cytotoxin, ExoU. *EMBO J* **22**, 2959–2969, <https://doi.org/10.1093/emboj/cdg290> (2003).
54. Yahr, T. L., Goranson, J. & Frank, D. W. Exoenzyme S of *Pseudomonas aeruginosa* is secreted by a type III pathway. *Mol Microbiol* **22**, 991–1003 (1996).
55. Yamaguchi, S. *et al.* Genotypic analysis of *Pseudomonas aeruginosa* isolated from ocular infection. *J Infect Chemother* **20**, 407–411, <https://doi.org/10.1016/j.jiac.2014.02.007> (2014).
56. de Almeida Silva, K. C. F. *et al.* Molecular characterization of multidrug-resistant (MDR) *Pseudomonas aeruginosa* isolated in a burn center. *Burns* **43**, 137–143, <https://doi.org/10.1016/j.burns.2016.07.002> (2017).
57. Georgescu, M. *et al.* Virulence and resistance features of *Pseudomonas aeruginosa* strains isolated from chronic leg ulcers. *BMC Infect Dis* **16**(Suppl 1), 92, <https://doi.org/10.1186/s12879-016-1396-3> (2016).
58. Tingpej, P. *et al.* Phenotypic characterization of clonal and nonclonal *Pseudomonas aeruginosa* strains isolated from lungs of adults with cystic fibrosis. *J Clin Microbiol* **45**, 1697–1704, <https://doi.org/10.1128/JCM.02364-06> (2007).

59. Berthelot, P. *et al.* Genotypic and phenotypic analysis of type III secretion system in a cohort of *Pseudomonas aeruginosa* bacteremia isolates: evidence for a possible association between O serotypes and exo genes. *J Infect Dis* **188**, 512–518, <https://doi.org/10.1086/377000> (2003).
60. Kulasekara, B. R. *et al.* Acquisition and evolution of the exoU locus in *Pseudomonas aeruginosa*. *J Bacteriol* **188**, 4037–4050, <https://doi.org/10.1128/JB.02000-05> (2006).
61. Feltman, H. *et al.* Prevalence of type III secretion genes in clinical and environmental isolates of *Pseudomonas aeruginosa*. *Microbiology* **147**, 2659–2669, <https://doi.org/10.1099/00221287-147-10-2659> (2001).
62. Azimi, S. *et al.* Presence of exoY, exoS, exoU and exoT genes, antibiotic resistance and biofilm production among *Pseudomonas aeruginosa* isolates in Northwest Iran. *GMS Hyg Infect Control* **11**, Doc04, <https://doi.org/10.3205/dgkh000264> (2016).
63. Feldman, M. *et al.* Role of flagella in pathogenesis of *Pseudomonas aeruginosa* pulmonary infection. *Infect Immun* **66**, 43–51 (1998).
64. Wolfgang, M. C., Jyot, J., Goodman, A. L., Ramphal, R. & Lory, S. *Pseudomonas aeruginosa* regulates flagellin expression as part of a global response to airway fluid from cystic fibrosis patients. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6664–6668, <https://doi.org/10.1073/pnas.0307553101> (2004).
65. Gellatly, S. L. & Hancock, R. E. *Pseudomonas aeruginosa*: new insights into pathogenesis and host defenses. *Pathog Dis* **67**, 159–173, <https://doi.org/10.1111/2049-632X.12033> (2013).
66. Fleiszig, S. M. J., Arora, S. K., Van, R. & Ramphal, R. FlhA, a Component of the Flagellum Assembly Apparatus of *Pseudomonas aeruginosa*, Plays a Role in Internalization by Corneal Epithelial Cells. *Infect Immun* **69**, 4931–4937, <https://doi.org/10.1128/iai.69.8.4931-4937.2001> (2001).
67. Spencer, C. & Brown, H. A. Biochemical characterization of a *Pseudomonas aeruginosa* phospholipase D. *Biochemistry* **54**, 1208–1218, <https://doi.org/10.1021/bi501291t> (2015).
68. Wilderman, P. J., Vasil, A. I., Johnson, Z. & Vasil, M. L. Genetic and biochemical analyses of a eukaryotic-like phospholipase D of *Pseudomonas aeruginosa* suggest horizontal acquisition and a role for persistence in a chronic pulmonary infection model. *Mol Microbiol* **39**, 291–303 (2001).
69. McMorran, B. J., Merriman, M. E., Rombel, I. T. & Lamont, I. L. Characterisation of the pvdE gene which is required for pyoverdine synthesis in *Pseudomonas aeruginosa*. *Gene* **176**, 55–59, [https://doi.org/10.1016/0378-1119\(96\)00209-0](https://doi.org/10.1016/0378-1119(96)00209-0) (1996).
70. Meyer, J. M., Neely, A., Stintzi, A., Georges, C. & Holder, I. A. Pyoverdinin is essential for virulence of *Pseudomonas aeruginosa*. *Infect Immun* **64**, 518–523 (1996).
71. Okuda, J. *et al.* Complementation of the exoS gene in the pvdE pyoverdine synthesis gene-deficient mutant of *Pseudomonas aeruginosa* results in recovery of the pvdE gene-mediated penetration through the intestinal epithelial cell barrier but not the pvdE-mediated virulence in silkworms. *Journal of Infection and Chemotherapy* **18**, 332–340, <https://doi.org/10.1007/s10156-011-0340-0> (2012).
72. Clinical and Laboratory Standards Institute (CLSI). Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically; Approved Standard—Ninth Edition. *CLSI* **32** (2012).
73. Clinical and Laboratory Standards Institute (CLSI). Performance standards for antimicrobial susceptibility testing; twenty-second information supplement. *CLSI document M100-S22* **32** (2012).
74. European Committee on Antimicrobial Susceptibility Testing (EUCAST). Breakpoint tables for interpretation of MICs and zone diameters. Version 6.0: The European Committee on Antimicrobial Susceptibility Testing. [http://www.eucast.org/clinical\\_breakpoints/](http://www.eucast.org/clinical_breakpoints/) (2016).
75. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170> (2014).
76. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455–477, <https://doi.org/10.1089/cmb.2012.0021> (2012).
77. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069, <https://doi.org/10.1093/bioinformatics/btu153> (2014).
78. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**, e11147, <https://doi.org/10.1371/journal.pone.0011147> (2010).
79. Rissman, A. I. *et al.* Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics* **25**, 2071–2073, <https://doi.org/10.1093/bioinformatics/btp356> (2009).
80. Carver, T. *et al.* Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **24**, 2672–2676, <https://doi.org/10.1093/bioinformatics/btn529> (2008).
81. 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).
82. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**, W16–21, <https://doi.org/10.1093/nar/gkw387> (2016).
83. Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693, <https://doi.org/10.1093/bioinformatics/btv421> (2015).
84. Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* **15**, 524, <https://doi.org/10.1186/PREACCEPT-2573980311437212> (2014).
85. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359, <https://doi.org/10.1038/nmeth.1923> (2012).
86. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, <https://doi.org/10.1093/bioinformatics/btp352> (2009).
87. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92, <https://doi.org/10.4161/fly.19695> (2012).
88. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* **67**, 2640–2644, <https://doi.org/10.1093/jac/dks261> (2012).
89. Jia, B. *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* **45**, D566–d573, <https://doi.org/10.1093/nar/gkw1004> (2017).
90. Chen, L., Yang, J. & Yu, J. VFDB: a reference database for bacterial virulence factors. **33**, D325–328, <https://doi.org/10.1093/nar/gki008> (2005).
91. Alikhan, N. F., Petty, N. K., Ben Zakour, N. L. & Beatson, S. A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12**, 402, <https://doi.org/10.1186/1471-2164-12-402> (2011).

## Acknowledgements

The authors would like to acknowledge the Singapore Centre for Environmental Life Sciences Engineering (SCELS), whose research is supported by the National Research Foundation Singapore, Ministry of Education, Nanyang Technological University and National University of Singapore, under its Research Centre of Excellence Programme. Sequencing of DNA was carried out with the help of Daniela Moses and Stephan Schuster using the sequencing facilities at SCELS. We are also thankful to UNSW high performance computing facility KATANA for providing us cluster time for data analysis.

### Author Contributions

D.S. designed the study, performed the experiments, analysed the data and wrote the drafts of the article. A.K.V. supervised D.S. and edited the article. G.S.K. helped with computational analyses and design of experiments. S.A.R. contributed to the design and implementation of the research, to the analysis of the results and edited the article. M.W. devised the project, developed the theoretical framework, edited the article.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-34020-7>.

**Competing Interests:** The authors declare no competing interests.

**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) 2018