Estimation of pathogenic potential of an environmental *Pseudomonas aeruginosa* isolate using comparative genomics

Carola Berger
Leibniz Institute for Natural Products Research and Infection Biology
Christian Rückert
Center for Biotechnology CeBiTec-Bielefield
Jochen Blom
Justus Liebig Universitat Giessen
Korneel Rabaey
Universiteit Gent
Jörn Kalinowski
Center for Biotechnology CeBiTec-Bielefeld
Miriam A. Rosenbaum (miriam.rosenbaum@leibniz-hki.de)
Leibniz Institute for Natural Products Research and Infection Biology https://orcid.org/0000-0002-4566-8624

Research article

Keywords: comparative genomics, genomic island, pathogenic potential P. aeruginosa

Posted Date: February 3rd, 2020

DOI: https://doi.org/10.21203/rs.2.22505/v1

Abstract

Background

New strains of *Pseudomonas aeruginosa* are continuously being isolated and sequenced to increase the genomic accessibility of this important pathogen. This has led to the generation of an impressive dataset of closed *P. aeruginosa* genomes. To understand the difference between the strains, investigations are focused on the accessory genome, thereby constantly extending the known pan genome of *P. aeruginosa* as a species. Apart from follow-up studies, many of the publicly available genomes are only used in their original publication while additional *in silico* information, based on comparison to previously published genomes, is not being explored. In this study, we defined and investigated the genome of the environmental isolate *P. aeruginosa* KRP1 and compared it to more than 100 publicly available closed *P. aeruginosa* genomes.

Results

Pseudomonas spp. KRP1 could clearly be identified as a *P. aeruginosa* isolate, via comparative genomics. By using different genomic island prediction programs, we could identify a total of 25 genomic islands that cover ~12% of the genome of *P. aeruginosa* KRP1. Based on intra-strain comparisons, we are able to predict the pathogenic potential of this environmental isolate. It shares a substantial amount of its genomic information with the highly virulent PSE9 and LESB58 strains. For both of these clones, their increased virulence has been directly linked to their accessory genome before.

Conclusions

Here we show, how the integrated use of previously published genomic data today, can help to replace expensive and time consuming wetlab work to determine the pathogenetic potential of environmental isolates. This knowledge is vital to understand what makes an isolate a potential pathogen as it helps design effective treatment.

Background

Pseudomonas aeruginosa is ubiquitously distributed. It has been isolated from terrestrial- and marine soil, fresh- and salt water, sewage, plants, animals and humans [1]. For the latter habitats, *P. aeruginosa* is known to act as an opportunistic pathogen that can infect plant- as well as animal models [2]. In humans, *P. aeruginosa* usually spreads to already vulnerable patients, like those with cystic fibrosis or who are hospitalized in intensive care units. About 10% of all nosocomial infections in most European Union hospitals are caused by this organism [3]. Its virulence is transmitted through the action of a myriad of virulence factors. An interesting facet of the virulence of *P. aeruginosa* is its combinatory aspect. This means, not every *P. aeruginosa* isolate and strain conveys an equal level of virulence to a given infection model and a strain that is effective in infecting a plant model does not necessarily show an equal amount of virulence towards an animal model [2, 4]. For the frequently researched *P. aeruginosa* PA14

strain this increased virulence, as compared to the type strain PAO1, is mainly due to the presence of additional virulence factors, whose genomic sequences are not part of the PAO1 genome. In PA14, most of those virulence factor encoding genes are clustered on two genomic islands (GIs) termed *P. aeruginosa* pathogenicity islands (PAPI) [5].

As in all living things, evolution leads to a change in the genomic composition of the bacteria's chromosome. Since their generation times are short, changes in the genomes are frequently observed and lead to the conclusion that bacterial genomes are viewed as dynamic rather than static gene collections [6]. For *P. aeruginosa,* numerous studies have proven that the pan genome can be viewed as a mosaic of a conserved core (~90% of the total genome) and variable accessory sections. Core genes are defined as genes with orthologues in nearly all strains of *P. aeruginosa* [7]. The core is marked by a conserved synteny of genes and a low average nucleotide substitution rate. The accessory genome, i.e. genes only shared by certain strains of a species, is mainly composed of GIs. They are clone- or strain-specific and lead to alterations in the genome size, which have been reported to range from 5.2 to 7.4 Mb in the *P. aeruginosa* species [2, 8]. By prokaryotic standards, this genome size is considered rather big. Especially since it is not a result of gene duplication events like for other microorganisms (e.g., *Neisseria meningitidis* [9]). As it encodes genes from numerous and distinct gene families, it highlights the great genetic and functional diversity of *P. aeruginosa* as a species [7]. A newer study has suggested the core genome of the *P. aeruginosa* species to consist of 4000-5000 open reading frames (ORFs) [2]. These genes make up the smallest portion of the pan genome. The second portion is the accessory genome of about 10,000 genes. These are genes found in multiple but not all strains. The by far largest fractions of the pan genome are singletons and rare genes that are only shared by very few strains. Their estimated number is at least 30,000 for the *P. aeruginosa* species [2].

Depending on the encoded genes, GIs can be classified into four functional categories: (i) pathogenicity islands (PIs), (ii) resistance islands (RIs), (iii) metabolic islands (MIs), and (iv) symbiotic islands (SIs). They can be also classified depending on general features like their means of inter- and intrachromosomal relocation. By assigning different functional modules, they can be grouped into (i) integrative and conjugative elements (ICEs), (ii) replacement islands, (iii) prophages and phage-like elements, and (iv) transposons, insertion sequences (ISs) and integrons [7].

Besides PAPI-I and PAPI-II of *P. aeruginosa* PA14, 42 other GI have been previously described in the *P. aeruginosa* species [10-13] of which multiple have been directly linked to an increased pathogenicity of the harboring strains [13-16]. Different detection software packages are available to help identifying regions of foreign DNA within a given genome. As the algorithms use different characteristics for this identification, usually not one program is able to identify all GIs. Hence, a combination of multiple complementary tools should be applied to get a thorough detection.

In this study, we describe how the abundantly available sequencing information of a species like *P. aeruginosa* can be used to characterize a newly sequenced strain. To this end, we sequenced the KRP1 environmental isolate and characterized its phylogenetic relationship by using more than 100 previously

published *P. aeruginosa* genomes. We further employed different GI detection software programs and manual mining to investigate the genome composition of this exemplary strain. By comparing the genomic content with other highly virulent *P. aeruginosa* variants, we are able to make educated predictions of the strains pathogenetic potential, without having to perform unnecessary and time- as well as money consuming animal experiments.

Results

Pseudomonas sp. KRP1 belongs to the *P. aeruginosa* species

In silico assembly of the *de novo* sequenced *Pseudomonas* sp. KRP1 resulted in two circular contigs of 6,162,740 bps and 575,136 bps, featuring a G+C content of 66.41% and 64.79%, respectively. Synteny comparisons between the initial *in silico* assembly and closely related *P. aeruginosa* strains showed multiple rearrangements of the ORFs encoded on the mega plasmid. In *P. aeruginosa* PA14, the sequence inscribed on the 575,136 bps contig are located between two homologous large ribosomal RNA clusters. These clusters are known to be spots of inner genome rearrangements within the *P. aeruginosa* species [4, 17]. The KRP1 homologue to PA14_61200 was also part of the mega plasmid and origin of a second reorganization event between PA14 and KRP1. As this large gene contains eleven 243 bp long tandem repeats, it has been found hard to be resolved by whole genome sequencing of a *P. aeruginosa* strain before [18]. Therefore PCR was used to investigate the DNA sequence surrounding the ribosomal RNA clusters on the main chromosome and on the potential mega plasmid. This resulted in a redefined genome structure of KRP1, with one circular chromosome, containing 6,301 protein-coding genes. The smaller contig was manually integrated at its correct position within the *in silico* sequence.

To clarify which species of the *Pseudomonas* genus KRP1 belongs to, the average nucleotide identity (ANI) percentage was calculated with respect to 105 fully sequenced *P. aeruginosa* strains and 8 other *Pseudomonas* species. When compared to the *P. aeruginosa* species, all calculated values are well above the accepted species threshold of 95-96% (Figure 1, for exact values see Table S1). For the eight other closely related *Pseudomonas* species, ANI values range between 80.4% (*P. citronellolis* P3B5) and 74.4% (*P. psychrotolerans* PRS08) (for all exact values see Table S1). While ANI comparisons are based on the nucleotide sequence, *P. aeruginosa* KRP1 was also analyzed based on the amino acid sequences of the transcribed proteins, towards the other strains and species. The tree was built based on a core of 1,537 genes per genome comprising 532,537 amino acid-residues per genome (Figure S1). For better visualization, a reduced version of the tree containing only the eight non-*aeruginosa* species and six *P. aeruginosa* strains is shown (Figure 2). The phylogenetic analysis clearly mark the strain KRP1 as a representative of the species *P. aeruginosa* and show a clear distinction of the strain towards other members of the same genus.

P. aeruginosa KRP1 relation to closely related *P. aeruginosa* strains

The phylogenetic trees in Figure 2 and Figure S1 are based on amino acid-sequences, and therefore present only non-synonymous nucleotide substitutions. For a more in depth investigation of KRP1, its genome was compared to the type strain PAO1, the highly researched strain PA14 and the two strains FA-HZ1 and W45909, to which KRP1 clusters most closely in the phylogenetic analyses (Table 1).

Table 1: Genomic overview of different *P. aeruginosa* strains used in this study.

P. aeruginosa strain	Total length [bp]	GC-content [%]	number of predicted genes	ANI with KRP1 [%]	Comment	Reference
KRP1	6,737,396	66.3	6,301			This study
PAO1	6,264,404	66.6	5,700	99.24	type strain	[17]
PA14	6,537,648	66.3	6,177	98.36	common research strain	[4]
LESB58	6,601,757	66.3	6,135	98.81	hyper virulent strain	[16]
FA-HZ1	6,866,790	66.2	6,389	99.98	closest sequenced relative to KRP1	[21]
W45909	6,777,566	66.2	6,475	99.96	2nd closest sequenced relative to KRP1	[22]

Regarding the synteny between KRP1 and the other *P. aeruginosa* strains, it becomes apparent that the type strain *P. aeruginosa* PAO1 (AE004091) shows a large-scale inversion with respect to the other five *P. aeruginosa* strains tested, which roughly affects 70% of the genome (Figure 3). This inversion of PAO1 with respect to other *P. aeruginosa* strains was observed before and is linked to an intra-chromosomal rearrangement at two of the four copies of a large ribosomal RNA cluster [4, 17]. When looking at the overall genome arrangement, KRP1 shows a high degree of synteny throughout the whole genome with the strains FA-HZ1, W45909 and PA14, which is not limited to certain parts.

The genome of *P. aeruginosa* has a mosaic-like structure, built of a conserved core (~ 90% of the total genome), which is interrupted by genomic islands containing variable accessory genes [7]. The numerical distribution between genes belonging to the core- and the accessory genome of the six *P. aeruginosa* strains (KRP1, PAO1, PA14, LESB58, FA-HZ1 & W45909) was analyzed using EDGAR and the results are summarized in Figure 4. The core genome shared by KRP1 and the two predominantly researched strains PAO1 and PA14 consists of 5,278 genes (Figure 4 A). This is equivalent to 83.8% (KRP1) – 92.6% (PAO1) of all genes annotated in the respective genomes (Table 1). KRP1 and PA14 contain orthologues of 216 genes, which are not part of the genome of the species type strain PAO1 (Figure 4 A, Area I). Additionally, there are 583 genes in KRP1, for which orthologues are not found in either of the two other strains (Figure 4 A, Area II). Thus, the environmental isolate KRP1 encodes for a substantially higher number of singletons than PAO1 or PA14. The overlap in the accessory genome of KRP1 is more pronounced with

the FA-HZ1 and W45909 strains of *P. aeruginosa* (Figure 4 B, Area III). Both strains also cluster as the closest relatives of KRP1 when only the core genome is taken into the calculation (Figure 2 & Figure S1). The three strains share a total of 5667 genes, which corresponds to 89.94% of all KRP1 predicted ORFs. The clinical isolate W45909 and the environmental isolate KRP1 both contain five genes related to inorganic ion transport and metabolism, which are not present in any of the other analyzed strains. All of these genes encode proteins related to copper metabolism, and might relate to common environmental challenges faced by the strains in their natural habitat. With the highly virulent LESB58 strain, KRP1 shares a total of 5503 (Figure 4 B, core + Area IV and V). In an inter-species comparison of these four strains (LESB58, FA-HZ1, KRP1 & W45909; Figure 4 B), the KRP1 genome encodes the lowest number of singletons (Figure 4 B, Area VI). Of these 102 genes, ~78% did not yield a BLAST hit within the COG database, highlighting that most of the genes of this area are novel or hypothetical proteins (Figure 5, Table S2). This high portion of unclassified genes was typical for all closer investigated overlap areas, except for the overlap of the KRP1 strain with LESB58 and W45909 (Figure 4 B, Area IV) . Here, the majority of the genes have a metabolic function and ~27% are related to cellular processes and signaling (Figure 5, Table S3).

The accessory genome of *P. aeruginosa* KRP1

In *P. aeruginosa* genes of the accessory genome tend to cluster into GIs [7]. Therefore, the genome of KRP1 was analyzed to detect putative genomic islands. Since the existing software programs have a different degree of sensitivity and different shortcomings, multiple programs were used: SIGI-HMM [23], IslandPath-DIMOB [24], PHASTER [25] and GIPSy [26]. The screening was complemented by a manual mining for known *P. aeruginosa* GIs. A list of the initial 28 GIs can be found in Klockgether et al. [10]. This list was extended by the findings of Silveira et al. [11], Hong et al. [12] and Jani et al. [13]. Over the years, a different nomenclature was established naming the islands PAPI-X (*P. aeruginosa* pathogenicity island), PAGI-X (*P. aeruginosa* genomic island) and LESGI-X (Liverpool Epidemic Strain genomic island). It is important to note that no direct correlation between PAGI and LESGI exists and that the respective islands are not exclusive to the PA or LES strains of *P. aeruginosa*. With the approach used here, a total of 25 putative GIs scattered throughout the KRP1 genome were detected (Table 2, Figure 6). The loci of the majority of these GIs could be clearly assigned to specific “regions of genomic plasticity (RGPs)” [27] (Table 2), which mark locations where integration of foreign DNA into the *P. aeruginosa* genome have been previously reported to happen with increased frequency. In general, GIs can be classified to a proposed metabolic function based on the encoded genes. The GIPSy software [26] categorizes GIs into i) pathogenicity islands (PIs), which contain an increased number of genes related to pathogenicity factors, ii) resistance islands (RIs) for genes related to antibiotic resistance, iii) metabolic islands (MIs) with genes known to be related to the biosynthesis of (secondary) metabolites, iv) symbiotic islands (SIs), which mainly contain genes related to a host-bacterium symbiotic relationship. In *P. aeruginosa* KRP1 all of these GI-classes are found, except for MIs (Table 2 & Figure 6) and some GIs are placed in more than one category. For a classification of a GI into the mentioned categories, it is not necessary that each single gene of the respective GI falls into the respective category.

The genome of KRP1 was also analyzed to identify which version of known replacement islands are encoded, as these traits represent critical determinants for the fitness and virulence of an individual *P. aeruginosa* strain [7] (Table 3). The four replacement islands contain the same functional content and occupy nearly always the same genomic loci within the *P. aeruginosa* core genome. Intriguingly, the specific genetic sequence of each island is highly diverse between strains [28, 29] The gene loci of the O-antigen gene cluster and the flagellin glycosylation replacement island are part of the PI/SI 5 and the RI/SI 11, respectively. The genes encoding pyoverdine production and pilin/pilin modification are not identified by the different genomic island detection programs. The pyoverdine locus is located between PI/RI 12 and GI 13, while the pilin modification genes are situated between PI 19 and GI 20.

Of the GIs recognized by the prediction software packages, PI/RI 1, GI 3, PI/RI 12 and GI 17 share a large portion of their nucleotide sequence with the other investigated *P. aeruginosa* genomes (i.e. with PA14: 50%, 80%, 80% and 90%, respectively). On the other hand, unique putative genes within them are assigned to only one of the analyzed strains and their integration into the core genome could be traced to a specific known RGP (Table 2). This classifies them as valid regions of the accessory genome of *P. aeruginosa.*

As GIs are usually elements of foreign DNA and frequently represent phage genomes, which have integrated into the host genome. The PHASTER software [25] and its predecessor PHAST [30] are designed to rapidly identify and annotate prophage sequences within a bacterial query genome. The software detected seven prophages throughout the KRP1 genome (Table S4). All of the detected sequences can be assigned to specific GIs and were therefore also recognized by the other genomic island detection programs tested. PHASTER classified four out of these seven prophages as intact, hence their genome contains all the necessary parts to be a complete phage and therefore to leave the genome again at a given point in time. Incomplete phages have likely lost essential parts of their genome, binding them permanently within the hosts DNA. One of these intact prophages is part of GI 23. The majority of the 51 putative ORFs show significant sequence similarities with the bacteriophage JBD93 [31] (92% identity over 86% of the query length) (Table S4). A highly similar prophage is integrated into the genome of the *P. aeruginosa* LESB58 strain, termed LES-prophage 4 [16]. The intact prophage genome is flipped within KRP1 and its ORFs are clustered into genes encoding proteins for tail assembly, head assembly facilitating genes, genes used for integration, and genes responsible for lysogeny and lysis. GI 23 also marks the only GI, for which no homolog could be detected in any of the other five investigated *P. aeruginosa* strains and it contains most of the detected singleton genes of KRP1 (Figure 6). Its integration disrupts the MdlC benzoylformate decarboxylase locus (PA14_64770), which has not been recognized as a RGP in *P. aeruginosa* before.

Table 2: Summary of genomic islands predictions in *P. aeruginosa* KRP1.

Genomic Island	Start position [bp]	Stop position [bp]	Size [bp]	KRP1 locus tag (number of ORFs)	RGP*	Prediction Method
PI/RI 1	40,389	61,808	21,419	KRP1_00205 - KRP1_00235 (7)	RGP46	2 & 4
GI 2	285,777	298,203	12,426	KRP1_01295 – KRP1_01335 (9)	RGP2	5
GI 3	671,911	697,058	25,147	KRP1_03145 - KRP1_03300 (32)	RGP3/4	1 & 3
PI 4	1,063,975	1,085,974	21,999	KRP1_05045 - KRP1_05145 (21)	RGP88	1, 2, 4 & 5
GI 5	1,222,896	1,230,252	7,356	KRP1_05780 – KRP1_05780 (1)	RGP89	5
GI 6	1,302,346	1,320,820	18,474	KRP1_06140 - KRP1_06225 (17)	RGP36	1 & 2
PI/SI 7	1,973,098	1,991,464	18,366	KRP1_09255 - KRP1_09325 (54)	RGP31	2 & 4
PI 8	2,424,758	2,470,270	45,512	KRP1_11590 - KRP1_11760 (36)	RGP28	1, 2, 3, 4 & 5
GI 9	2,533,287	2,538,355	5,068	KRP1_12025 – KRP1_12060 (8)	-	2 & 5
GI 10	2,556,402	2,564,724	8,322	KRP1_12155 – KRP1_12190 (8)	RGP71	5
PI/RI/SI 11a	2,632,036	2,744,677	112,641	KRP1_12500 – KRP1_13040 (109)	RGP27	1, 2, 4 & 5
GI 11b	2,751,082	2,753,517	2,435	KRP1_13080 – KRP1_13095 (4)	RGP27	2
PI/RI 12	2,895,779	2,921,721	25,942	KRP1_13740 – KRP1_13765 (7)	RGP25	4 & 5
GI 13	3,221,391	3,272,809	51,418	KRP1_14895 – KRP1_15110 (44)	RGP23	1, 2, 4 & 5
GI 14	3,577,280	3,579,282	2,002	KRP1_16510 – KRP1_16510 (1)	RGP52	5
GI 15	3,769,299	3,777,692	8,393	KRP1_17360 – KRP1_17400 (9)	-	5
RI/SI 16	4,485,821	4,496,553	10,732	KRP1_20830 – KRP1_20870 (9)	RGP9	4
GI 17	4,592,095	4,616,393	24,298	KRP1_21355 – KRP1_21485 (27)	RGP7	1 & 5
GI 18	4,762,338	4,768,531	6,193	KRP1_2211 – KRP1_22245 (8)	RGP6	5
PI 19a	4,867,542	4,906,902	39,360	KRP1_22720 – KRP1_22960 (49)	RGP5	1, 3 & 4
PI 19b1	4,906,929	4,925,297	18,368	KRP1_22965 – KRP1_23060 (20)	RGP5/41	1 & 4
PI 19c	4,925,522	4,955,315	29,793	KRP1_23065 – KRP1_23155 (19)	RGP41	2 & 4
PI 19b2	4,955,299	4,983,156	27,857	KRP1_23160 – KRP1_23310 (31)	RGP5/41	1, 2 & 4
PI 19d	4,983,197	5,009,461	26,264	KRP1_23315 – KRP1_23425 (23)	RGP5	1, 2, 3 & 4
GI 20a	5,366,804	5,428,778	61,975	KRP1_25040 – KRP1_25385 (70)	RGP41	1, 2 ,3, 4 & 5
GI 20b	5,455,015	5,464,821	9,807	KRP1_25540 – KRP1_25565 (6)	RGP41	2, 4 & 5
GI 21	5,615,479	5,626,409	10,930	KRP1_26250 – KRP1_26310 (13)	-	5
GI 22	5,700,164	5,727,413	27,249	KRP1_26660 – KRP1_26790 (7)	-	5
GI 23	5,875,381	5,911,730	36,349	KRP1_27515 – KRP1_27765 (51)	-	1, 3, 4, 5
GI 24a	6,203,408	6,209,504	6,096	KRP1_29015 – KRP1_29040 (6)	-	5
PI 24b	6,209,865	6,225,427	15,562	KRP1_29045 – KRP1_29090 (10)	RGP62	1, 2, 4 & 5
GI 24c	6,225,700	6,237,151	11,451	KRP1_29095 – KRP1_29150 (12)	-	5
PI 24d	6,239,438	6,281,035	41,597	KRP1_29155 – KRP1_29380 (46)	RGP87	1, 3, 4 & 5
GI 24e	6,281,429	6,299,576	18,147	KRP1_29385 – KRP1_29450 (14)	-	5
GI 25	6,397,652	6,402,302	4,650	KRP1_29920 – KRP1_29930 (3)	-	2

* Reported regions of genomic plasticity - RGPs 1–62: [27]; RGPs 63–80: [33]; RGP 81-86: [16]; RGP 87-89: [34]; RGP 90-97: [35].

Prediction method: 1: IslandPath-DIMOB [24]; 2: SIGI-HMM [23]; 3: PHASTER [25]; 4: GIPSy [26]; 5: manual blast against previously described *P. aeruginosa* GIs.

Table 3: Replacement islands in *P. aeruginosa.*

Replacement island	subgroups	RGP*	PAO1	PA14	LESB58	FA-HZ1	W45909	KRP1
O-antigen biosynthetic locus	20 [36]	RGP31	O5 [37]	O10 [4]	O6 [38]	O1 (this study)	O1 (this study)	O1 (this study)
Pyoverdine locus	3 [39]	RGP73	type I [28]	type I [28]	type III [13]	type I (this study)	type I (this study)	type I (this study)
Pilin and pilin modification genes	5 [40]	RGP60	group II [40]	group III [40]	group I [41]	group I (this study)	group I (this study)	group I (this study)
Flagellin glycosylation island	2 [42]	RGP9	b-type [42]	b-type [43]	b-type [44]	a-type (this study)	a-type (this study)	a-type (this study)

* RGPs 1–62: [27]; RGPs 63–80: [33].

Besides GI 23, PI 24d is a predicted intact prophage (Table S4). The same φCTX phage was previously detected within the *P. aeruginosa* PSE9 strain [34], where it was termed PAGI-6. Interestingly, the φCTX phage integrated itself more than once into the KRP1 genome. It is also encoded within the first part of PI 19. In both loci, the prophage is lacking the φCTX cytotoxin producing gene *ctx*. While the PI 24d copy of the prophage does contain two pseudogenes at the ctx position, the PI 19a copy does not contain this 1,589 bp stretch of DNA. Hence, this loci marks a RGP of the prophage and further highlights that the *ctx* gene is not part of the ancestral φCTX genome but rather was incorporated by horizontal gene transfer later on, as has been previously suggested in literature [31, 45, 46]. A second RGP within the φCTX genome of PI 19a, spans the 5 ORFs KRP1_22865 - KRP1_22885 – encoding for proteins of unknown functions, which differ from PAGI-6 and the known φCTX genome. Another difference between PI 24d and PI 19a is the encoding of two recombinases (KRP1_29355 & KRP1_29375) and an integrase (KRP1_29370) only in PI 24d, while PI 19a, encodes a different integrase (KRP1_22960), which shows sequence similarities with KRP1_12505, the integrase encoded at the start of PI/RI/SI 11a. This is little surprising, since both integrases facilitate the integration downstream of a $tRNA^{Gly}$, while PI 24d is integrated downstream of a $tRNA^{Thr}$.

GI integration downstream of a tRNA is a well-studied phenomenon [47, 48]. The 3'-ends of tRNAs carry *attB* sites, which are recognized and used for site-specific recombination between an integrative and conjugative element (ICE) and the main chromosome. Overall, the integration of PI 8, 19 and 24b&d, 16 and 17 as well as GI 17, 11, and 20 occurred just downstream of specific tRNAs within the KRP1 genome. Of these islands, GI 17, PI 19 and GI 20 belong to the same family of *P. aeruginosa* GIs, which are marked by their bipartite structure. While the first segment, downstream of the tRNA contains strain-specific cargo ORFs, the second part shows a high degree of sequence similarity between the strains [16, 47] and mainly encodes structural and mobility-related genes, as well as genes for conjugal transfer [10]. Due to their conserved structure, islands of this family and similar islands in other β- and γ-proteobacteria are likely to be derived from one common ancient ancestor [49]. Additional, previously detected and analyzed GIs of this family include PAGI-2, PAGI-3 and LESGI-3 [16, 47]. Cargo genes of GI 8 include heavy metal resistance genes, genes for metabolic enzymes and enzymes used for the formation and altering of nucleic acids, transcription regulators, a two-component system as well as an antibiotic resistance gene. While the here analyzed cargo genes are KRP1 specific with respect to PAGI-2, PAGI-3 and LESGI-3, they share 99% sequence identity with 13 of the 105 *P. aeruginosa* isolates, used for phylogenetic comparison (Table S1). These include the previously mentioned FA-HZ1 and W45909 strains. We hypothesize that this set of cargo genes form a unit, which contributes to the successful survival of *P. aeruginosa* in certain habitats.

PI 19 is the structurally most complex genomic island of KRP1. PI 19b shows substantial sequence similarities to the second half of GI 11 and to LESGI-3 [16] (Figure 7). PI 19b mainly encodes proteins that relate to bacterial conjugation including part of a *tra*-like operon (KRP1_22995, KRP1_23010 & KRP1_23030), known to facilitate plasmid transfer in gram negative bacteria [51]. Instead of an array of cargo genes, as in GI 11, PI 19 has the integrated prophage PI 19a (φCTX phage) upstream of the conserved block. The *tra*-like transfer operon of PI 19b is interrupted by another section of DNA described as PI 19c, which is likely of foreign origin. The G+C content of 62.32% in PI 19c differs noticeably from the surrounding PI 19b parts with G+C contents of 68.21% and 67.17%, respectively. The enclosed genes have their closest sequence homologues outside the *Pseudomonas* genus and include i) genes associated with the genomic repair system of the cell (DNA repair proteins (KRP1_23085 & KRP1_23105) and RNA polymerase-associated proteins, needed for RNA polymerase recycling (KRP1_23090 & KRP1_23110)) and ii) a cluster of metabolic enzymes (thiamine biosynthesis protein (KRP1_23140), a cyclic AMP-GMP synthase (KRP1_23145) and a patatin-like phospholipase (KRP1_23150)). Interestingly, no genes associated with relocalization of genomic DNA are encoded in this region of the island. This suggests that the integration into PI 19b occurred a long time ago, followed by partial deletion of the respective genes or its integration by a yet unknown mechanism. The *tra*-like transfer operon of PI 19b is continued at the KRP1_23175 ORF, highlighting the interruptive nature of PI 19c within PI 19b. The rest of PI 19b contains mainly genes for proteins of unknown function. The last 23 ORFs of PI 19 likely form an additional separate sub-section PI 19d, as they contain multiple elements related to transposable rearrangement of DNA (KRP1_23345 - KRP1_23360) and other phage related proteins (Table S4). Other cargo genes of this sub-island include multiple proteases (KRP1_23370, KRP1_23380, KRP1_23425 &

KRP1_23430) and genes linked to different stress responses like acid tolerance (KRP1_23390 & KRP1_23395) and phosphate starvation (KRP1_23420). This last part of PI 19 represents classical strain specific cargo genes, which are typical for the previously explained PAGI-2 and PAGI-3 like family of *P. aeruginosa* GIs [10].

Genomic resemblance of KRP1 to highly virulent *P. aeruginosa* strains

P. aeruginosa KRP1 contains an array of genomic elements that are found in the highly virulent strains PSE9 [14, 34] and LESB58 [13, 16, 52] (Table 4). Unfortunately, no complete genome sequence is available yet for PES9, so it could not be included in the full genome comparison. However, some of the shared GIs have been shown to be the source of the strain dependent virulence within the P. aeruginosa species [14-16]. KRP1 encodes all seven genomic islands found in the clinical isolate PSE9 [14, 34] (Table 4), whereby PAGI-9 (GI 5) and PAGI-11 (GI 14) were not detected by the GI prediction tools used in this study, but their presence within the KRP1 genome was manually verified (Table 2). PAGI-9 consists of 6581 bp and one large ORF, which was identified as a Rhs (rearrangement hot spot) element [34]. The nucleotide sequence of these proteins generally has a bipartite structure composed of a long G+C rich core and a relatively G+C poor tip sequence. While the core sequence is intra- and interspecies highly conserved, the tip is rather variable. Rhs elements in P. aeruginosa are often linked to- and translocated via type VI secretion systems [53-55]. Other members of the Rhs element family have been shown to exhibit bacteriocin properties, highlighting their use in inter-prokaryotic competition [56]. Similarly, PAGI-10 is a Rhs element of PSE9, which is also found within KRP1 (PI/RI 9). The fact that the strains PSE9 and KRP1 show sequence identity over the entire length of the ORF and not only in the conserved core shows the close genomic relationship between the hyper virulent PSE9 and KRP1.

PAGI-11 of PSE9 (GI 14 in KRP1) is only 2,003 bp long and located at RGP 52 (Table 4) and while Battle et al. [34] did not find any ORFs contained, the Prokka pipeline [57], used on the KRP1 genome, predicts the hypothetical protein KRP1_16515. The G+C content of just 43.19% is far below the average of the KRP1 genome (i.e. 66.3%). Other strains are known to contain GIs of larger size and with encoded mobile element related genes at this specific genomic locus [27]. Therefore, PAGI-11 might have been a larger genomic island in the past, which was partially lost over time in PSE9 and KRP1. The PSE9 strain originated from a patient with ventilator-associated pneumonia isolated at a hospital in Barcelona, Spain, between May 1993 and October 1997 [58]. It was found to be the most virulent out of 35 strains in a mouse model of acute pneumonia [59]. So far, two studies were able to link the increased virulence of PSE9 directly to PAGI-5 and PAGI-9 [14, 15]. Since KRP1 contains both of the mentioned islands, an increased virulence similar to the levels of PSE9 can be anticipated. Further, PSE9 and KRP1 share the same O-antigen type O1 (Table 3). The O-antigen type of the outer membrane lipopolysaccharide (LPS) layer has been previously linked to the virulence of *P. aeruginosa*, but most studies consider the serotype of the type strain PAO1 (type O5) for their study [60]. Both strains are also *exoS* positive and *exoU*

negative. A genotype that has been linked to an invasive phenotype [61]. Since no full genome sequence of PSE9 is available so far, a deeper *in silico* comparison between both strains is currently impossible.

Besides PSE9, the *P. aeruginosa* strain KRP1 shows substantial similarities in its accessory genome with the LESB58 strain, an aggressive pathogen of a cystic fibrosis patient from Liverpool in 1988 [13, 16, 52] (Table 4). The majority of the shared GIs were found via manual search rather than by the applied software programs. LESGI-6 to LESGI-17 were first detected by Jani et al. [13]. The authors used a genome segmentation approach to identify genomic regions of foreign origin within the LESB58 strain. This technique varies from the ones used in this study and therefore different putative GIs were detected. The authors could show that these GI encode for additional virulence factors (LESGI-6, -8, -13, and -15) as well as drug and metal resistance cassettes (LESGI-12 and -17). LESGI-9, -16, and -17 add additional versatility to the LESB58 metabolic repertoire [13]. Since KRP1 encodes all of these GIs as well, it is very likely that it employs their functions and therefore shows an increased virulence potential, similar to the LESB58 strain.

Table 4: Genomic Islands in different *P. aeruginosa* strains. GIs of strain PSE9 and selected GIs of strain LESB58 and their corresponding GI as well as sequence similarity within the strain KRP1.

	Location within KRP1	Sequence identity (Query length)
PSE 9 GIs		
PAGI-5	GI 20	99.98% (98%)
PAGI-6	PI 24d	99.98% (100%)
PAGI-7	PI 4	100% (100%)
PAGI-8	PI 24b	99.99% (100%)
PAGI-9	GI 5	100% (100%)
PAGI-10	PI/RI 9	99.97% (100%)
PAGI-11	GI 14	100% (100%)
LESB58 GIs		
LESGI-1	PI 8	98.62% (96%)
LESGI-3	GI 11	99.54% (65%)
LESGI-4	GI 13	99.61% (98%)
LESGI-6	GI 2	99.36% (100%)
LESGI-8	GI 9	99.41% (100%)
LESGI-9	GI 10	99.75% (100%)
LESGI-12	GI 15	99.55% (100%)
LESGI-13	GI 17	99.60% (100%)
LESGI-14	GI 18	99.56% (100%)
LESGI-15	GI 21	99.73% (100%)
LESGI-16	GI 22	99.62% (100%)
LESGI-17	GI 24	99.59% (96%)
LES-prophage 4	GI 23	89.31% (73%)

In contrast, the two strains showing the closest ANI identity and phylogenetic relationship with KRP1 are *P. aeruginosa* strain FA-HZ1 and W45909 (Figure 1 and Figure 2). FA-HZ1 is a dibenzofuran-degrading isolate from China [21] while W45909 is a clinical isolate from the USA [22]. All but three identified GIs in KRP1 are also present in these two most related strains (PI 8, PI 19 & GI 23 for W45909 and GI 23 for FA-HZ1). This provides circumstantial evidence that the genomic repertoire of *P. aeruginosa* KRP1 is likely to sustain a pathogenic as well as an apathogenic lifestyle in nature. While their genetic information is

available, no further studies have been performed with either of these strains but we stand to believe that they will also show an increased virulence like KRP1.

Discussion

Pseudomonas sp. strain KRP1 was first isolated from a microbial fuel cell as one of the dominating bacterial species responsible for high electron transfer efficiency of the mixed community [62]. In this original study the highest similarity BLAST hit of 95% similarity was found with *Pseudomonas aeruginosa* ATCC 27853 using a 197 bp fragment of the 16S rRNA gene. This study now clearly identified KRP1 as a strain variant of the species *P. aeruginosa* using a core genome comparison approach against 105 fully sequenced *P. aeruginosa* strains and 8 other *Pseudomonas* species (Figure S1). Sequence differences within the *P. aeruginosa* clade are very small when taking only the core genes (i.e. genes with orthologues in all analyzed strains) into consideration (Figure S1), i.e. earlier studies found the average rate of sequence polymorphism to be 0.3% among *P. aeruginosa* housekeeping genes [63] and an interclonal sequence diversity of 0.5–0.7% within core components [64, 65]. For the two closest genetic relatives, the strains FA-HZ1 and W45909 only their sample origins and genomes are known so far. FA-HZ1 is an environmental isolate from China, which was characterized for its dibenzofuran-degrading ability [21], while W45909 is a clinical isolate from the USA [22].

Due to their close phylogenetic relationships, the strains FA-HZ1 and W45909 were chosen for an in depth genomic comparison with *P. aeruginosa* KRP1. Additionally, the type strain PAO1, the well-researched strain PA14, and the highly virulent LESB58 were included in this multi way comparison (Table 1). For *P. aeruginosa* numerous studies have proven that the genome can be viewed as a mosaic of a conserved core and variable accessory sections [2, 7]. Hilker et al. [2] have suggested that the core genome of *P. aeruginosa* consist of 4,000 – 5,000 ORFs. It is therefore the smallest portion of the *P. aeruginosa* pan genome. For the six strains investigated in this study, the core is comprised of 4978 genes, which corresponds to 76.9% (W45909) – 87.3% (PAO1) of all genes annotated in the respective genomes (79% for KRP1). These values are in good agreement with literature values that report the core to make up ~ 90% of the total genome [7]. The core is marked by a conserved synteny of genes and a low average nucleotide substitution rate. It predominantly includes primary metabolism related genes, as well as genes involved in transcription and translation [66].

As a species, *P. aeruginosa* contains another 10,000 genes, which make up the accessory genome. For the strains KRP1, FA-HZ1, and W45909 a large part of their accessory genome is overlapping (area III Figure 4), accumulating to 89.94% total overlap (core + common accessory genes) for KRP1. This is interesting, since all three strains originate from three different habitats and continents. This combination of core and accessory genes seems to enable the strains to thrive in a pathogenic (W45909) as well as an environmental (KRP1 and FA-HZ1) habitat. ~70% of all accessory genes shared between the three strains are only poorly described or yielded no hits against the COG database. With the potent pathogen LESB58, KRP1 shares 299 accessory genes. The majority of classified genes belong to the subgroup of cellular

processes and signaling subgroup or are related to metabolism, which gives a hint that the biological niches occupied by these strains seems to be similar.

Singletons and rare genes (i.e., genes without any detectable orthologues in other strains) are by far the largest fraction of the *P. aeruginosa* pan genome. The strain KRP1 contains 65 singletons with respect to the other five strains. In the individual sub-set comparison this number is substantially higher, with 583 unique genes when compared to PAO1 and PA14 (area II Figure 4) and 102 singletons in comparison to LESB58, FA-HZ1 and W45909 (area VI Figure 4). The majority of them are not classified within the COG database, but are recognized as phage related proteins by the PHASTER software and are located within the identified GIs of KRP1.

The genes of the accessory genome, are not scattered randomly throughout the *P. aeruginosa* KRP1 genome, but are mainly clustered in 25 GIs, which in turn tend to integrate in certain genomic loci termed RGP [27] (Figure 6). Multiple GIs, whose possible presence in a *P. aeruginosa* genome is known from literature, were not detected by any of the used software tools. Their existence in KRP1 was determined via manual scanning of the genome. This highlights on the one hand, the usefulness of the multiple program approach for detection of putative genomic islands within a novel sequenced strain. On the other hand, it shows that the detection algorithms of the programs are not perfect and by just relying on them, relevant information might be overlooked. It is therefore crucial to complement the *in silico* analysis by implementing previously reported results to obtain a comprehensive view of the genomic structure of a newly sequenced strain. Since the overall average G+C content of *P. aeruginosa* KRP1 is at 66.3% (Table 1) and therefore considered GC rich, genes acquired through horizontal gene transfer usually have a lower G+C content (black ring in Figure 6). After integration of the foreign DNA into the chromosome, it is subject to the same selective evolutionary pressure as the rest of the host chromosome. Thus, over time it is likely to lose the sequence compositional differences making it undistinguishable from genomic material originating from *P. aeruginosa* [7]. These regions are therefore not detected by GI prediction software targeting differences in sequence composition. In the case of the 25 putative GIs in KRP1, most have a noticeable lower G+C content compared to the surrounding core genome and are therefore of rather young evolutionary origin. Several of the homologous PAGI and LESGI GIs in KRP1 in contrast were not detected by any of the used algorithms, which might point to an evolutionary older event of acquisition of these GIs.

The genome of KRP1 was also analyzed to identify, which version of known replacement islands are encoded (pilin/pilin modification, flagellin glycosylation island, O-antigen gene cluster and pyoverdine production) (Table 3). These four characteristic features are encoded in gene clusters, which contain horizontally acquired components. While they are highly divergent between different strains, the clusters occupy nearly always the same genomic loci within the *P. aeruginosa* genome [7]. As O-antigens, pili and flagella are recognized targets for phage entry and the host immune system, keeping different varieties of the same gene locus is thought to be a defense mechanism of *P. aeruginosa* [7]. Pyoverdines are an entry target for pyocins, bacterially produced phage-like molecules with antibacterial properties [67]. These

traits represent critical determinants for the fitness of an individual *P. aeruginosa* strain [7]. In detail, the replacement islands of KRP1 are characterized as follows:

(i) KRP1 expresses the most common pilin variant (group-I). Pilin is the monomeric subunit of the type IV pili in *P. aeruginosa*. Besides mediating adherence to host cell surfaces and involvement in biofilm formation, they are often used as receptors by *P. aeruginosa* phages [68-71]. In group-I, the pilin subunit is glycosylated at a C-terminal serine. This glycosylation rules out the pilus as receptor points for certain bacteriophages [72]. Group-I has been linked increasingly to cystic fibrosis environments [40].

(ii) Flagellins, the protein subunits of the polar flagellum, are classified into an a-type (identified for KRP1) or b-type, which is based on the molecular weight and the reactivity with type-specific antisera. They are glycosylated post-transcriptionally by the gene products of a type specific glycosylation island composed of 14 genes for the a-type [73] and four gens for the b-type [43]. While the different glycosylation patterns do not affect flagellum-mediated motility of the cells, they play a significant role in strain virulence [43, 74, 75].

(iii) Pyoverdine is a siderophore used by *P. aeruginosa* to accumulate iron(III). Based on the ligands bound to the inner structure of the protein, they can be grouped into three siderovars [76]. Each of these subgroups produces a specific combination of pyoverdine and its corresponding receptor [7]. Additionally to being a valuable trait in natural environments, pyoverdine production has been connected to virulence of *P. aeruginosa* in the past [77, 78]. It is the only replacement island in which KRP1, PAO1, and PA14 contain the identical variant.

(iv) The O-antigens are the outermost part of the LPS layer and consist of repeating sugar moieties. The gene cluster encodes proteins of a functional unit, which mediates the chemical modification and assembly of the sugar residuals, as well as their translocation to the periplasm. Also the ligation of these smaller lipopolysaccharides into repetitive chains is encoded in the gene cluster [7]. Research has been focused on the link between O-antigens and virulence [60, 79] but no specific results for the here found O1-serotype are known, as most studies focus on the O5-serotype of the type strains PAO1 [60]. Some phages depend on the O-antigen for infection. In the case of KRP1, the intact JBD93 bacteriophage, which was detected as GI 23, uses O-antigen mediated infection [31]. Since PAO1 and PA14 encode the genes for different O-antigen serotypes (Table 3), they are likely no targets for JBD93. Therefore, almost all of the 51 ORFs encoded in GI 23 are unique to KRP1 in the inter-strain comparison (Figure 4A). Even though, the closely related FA-HZ1 and W45909 strains also have the O1-serotype, the prophage is not encoded in their genome. Further, the integration point of GI 23 is not a previously recognized RGP. This leads us to believe that the prophage integration into the KRP1 genome is a recent evolutionary event. Besides JBD93, multiple other prophages have integrated into the KRP1 genome (Table S4).

The *P. aeruginosa* KRP1 isolate was so far investigated only towards its ability to produce increased levels of current in bioelectrochemical systems [80] and to dominate such systems, when inoculated with a mixed community of microbes [62]. Thereby, its main success for function in bioelectrochemical systems and likely also a factor in its dominance of microbial communities, is its enhanced production of

phenazines [80]. These redox active natural products mediate the electron transfer to a provided electrode and act as antimicrobials through the generation of reactive oxygen species. Phenazines are also considered important virulence factors during *P. aeruginosa* infections [81, 82]. Their production is encoded on two redundant operons, which are located within the core genome. The same holds true for many other important known virulence factors of *P. aeruginosa* [83], as no apathogenic variants of the species have been reported so far. However, a strong intraspecies gradient of virulence is observed in *P. aeruginosa*, ranging from highly infective to only mellow virulent strains [2, 14]. This phenomenon is likely linked to the varying accessory genome of the variants. Based on the genome analysis presented here, overall predictions of the virulence of KRP1 are possible. KRP1 contains all seven genomic islands found in the highly virulent *P. aeruginosa* clinical isolate PSE9 [14, 34] (Table 4). So far, two studies were able to link the increased virulence of PSE9 directly to PAGI-5 and PAGI-9 [14, 15]. As KRP1 contains both of the mentioned islands, an increased virulence similar to the levels of PSE9 can be anticipated. Further, PSE9 and KRP1 share the same O-antigen type O1 and are *exoS* positive and *exoU* negative. The type III exoenzyme and its cognate chaperone SpcS are encoded on GI 17. This genotype has been linked to an invasive phenotype [61]. Unfortunately, no full genome sequence of PSE9 is available so far, rendering a deeper *in silico* comparison between both strains impossible.

Besides PSE9, the *P. aeruginosa* strain KRP1 shows substantial homologies in its accessory genome with the aggressive pathogenic strain LESB58 strain [13, 16]. The strain is beta-lactam-resistant [52], exhibits enhanced survival on dry surfaces [84], shows an increased patient morbidity [85], and overexpression of parts of the quorum sensing regulon during early growth phases (e.g. LasA, elastase, and pyocyanin) [86, 87]. It is also known to replace previously established *P. aeruginosa* strains due to its aggressive nature, thereby causing a superinfection [88]. A LES isolate has even been reported to infected the non-CF parents of a CF patient [89]. While the complete reasons for its increased virulence are still partially unknown, naturally a lot of the responsible factors are thought to be driven by the accessory genome of the strain [13, 16]. These LESGI termed genomic islands differentiate the LES strain from other *P. aeruginosa* strains. Of the 17 known LESGIs and six LESGI-prophages, the genome of KRP1 contains 12 LESGIs and one prophage (Table 4). In a more recent study, Jani al. [13] could show that the shared GIs LESGI-6, 8, 13, and 15 encode for additional virulence factors, LESGI-12 and -17 enable additional drug and metal resistance, while LESGI-9, -16, and -17 expend the metabolic repertoire of LESB58 and KRP1, respectively. The close genomic relationship in the accessory genome of KRP1 with the two potent pathogens LESB58 as well as PSE9 gives us reason to believe that this environmental isolate also exhibits increased virulence.

Conclusions

The genome of the BES isolate *Pseudomonas* sp. KRP1 was *de novo* sequenced and analyzed in depth for its phylogenetic relationship within the *Pseudomonas* clade. Due to the sequence composition of its core genome, it could clearly be assigned to belong to the *P. aeruginosa* species. Its two closest relatives are two recently sequenced strains from China (FA-HZ1) [21] and the USA (W45909)[22].

The accessory genome of KRP1 was thoroughly analyzed. Using four different prediction programs, 25 putative genomic islands were detected. This analysis was extended by mining for the 44 GI complexes previously described in *P. aeruginosa* [10-13]. Most of the GIs could clearly be assigned to a known RGP (Table 2). The majority of the KRP1 singletons, with respect to the strains PAO1, PA14, FA-HZ-1, W45909 and LESB58, are contained in these islands, marking them as the main source of genome divergence between the strains.

Utilizing the increased amount of sequencing data, made publicly available in the past decade, it is possible to make *in silico* based educated prediction towards the virulence potential of a newly isolated strain of *P. aeruginosa*. Hence, it decreases the need for laborious wet lab experiments and animal testing. Using publicly available data and integrate them with own research data can help to substantially speed up research in the future and to draw wider, more general conclusions. The degree to which the individual GIs contribute to virulence of the strain is still to be determined and proves to be a rather difficult task since virulence in *P. aeruginosa* is known to be combinatorial [4, 90].

Methods

Strain and Medium

P. aeruginosa KRP1 was isolated from a microbial fuel cell setup at the Laboratory of Microbial Ecology and Technology (LabMET) at Ghent University (deposited into the Belgian Co-ordinated Collections of Microorganisms, BCCM; strain number LMG 23160) [62]. Cultures were grown in shake flasks in Luria Broth medium at 37°C, 200 rpm shaking,

DNA sequencing

Genomic DNA of *P. aeruginosa* KRP1 was isolated via phenol-chloroform extraction [91; mod.]. Besides a purity check on a NanoDrop™ One/OneC Microvolume-UV-Vis-Spectrophotometer (Thermo Fisher Scientific) and an integrity check via agarose gel electrophoresis, the concentration of isolated genomic DNA was estimated via a PicoGreen® dsDNA quantification assay (Quant-iT™ PicoGreen™ dsDNA Assay Kit, Thermo Fisher Scientific). The measurement for this assay was done with a Synergy Mx® microplate reader (BioTek) in 96-well plates using an excitation wavelength of 480 nm, an emission wavelength of 520 nm, a scan width of 9.0 and an overflow value of 80.

For shotgun library preparation 1 µg of chromosomal DNA was used (TruSeq DNA PCR-Free Library Preparation Kit, Illumina). Samples were sequenced on an Illumina MiSeq system using the MiSeq Reagent Kit v3 for 600 cycles. The data (542.3 Mb equaling ~ 81.3x coverage) were assembled using Newbler v.2.8 (Roche), which resulted in 58 scaffolds containing 94 contigs. Gap closure was conducted with a MinION Mk1B Sequencer from Oxford Nanopore Technologies. For this second shotgun library, 2 µg of genomic DNA was used as starting material. Size-selected DNA-fragments of 5 to 50 kb were used

to create a $1D^2$ sequencing library according to the manufacturer's instructions (1D² Sequencing Kit (R9.5), Oxford Nanopore Technologies). The sequencing library was run on a R9.5 flowcell for 3 hours. Base calling and data conversion to FastQ was performed using Albacore v1.2.4 [92https://github.com/Albacore/albacore]. The resulting 72.4 Mb (12x coverage) sequencing data were assembled with Canu v1.5 [93]. After assembly, the resulting 23 contigs were polished with the short Illumina reads using PILON [94]. The final assembly was done manually using Consed [95] to combine the contigs of the Newbler and Canu assemblies. Gene prediction and annotation of the finished genome were performed using the Prokka pipeline [57]. Visualization and inspection of the annotated sequence was done in Artemis [96].

To clarify the existence of a potential mega plasmid, a PCR using EconoTaq® PLUS GREEN DNA polymerase (Lucigen) was performed. The PCR fragments were sequenced by Eurofins Genomics (Germany).

Comparative genome analysis

For the analysis of the assembled KRP1 genome in respect to other *Pseudomonas* genomes and to find orthologous genes in related genomes, the EDGAR ("Efficient Database framework for comparative genome analyses using BLAST score Ratios") [19, 20] platform was used. For functional gene classification, ORFs were checked against the Clusters of Orthologous Groups (COG) database [97]. Parameters were set to an e-value of $<1e^{-10}$ and 80% identity. The genome of KRP1 was compared to 105 fully sequenced *P. aeruginosa* strains and 8 other *Pseudomonas* species. More in depth analysis were performed with the type strain PAO1 (AE004091; NC_002516.2; [17]), the frequently researched strain PA14 (UCBPP-PA14; NC_008463.1; [4]), the highly virulent strain LESB58 (NC_011770.1; [16]) and the two phylogenetically closest strains FA-HZ1 (NZ_CP017353.1; [21]) and W45909 (NZ_CP008871.2; [22])

Multiple genomic island prediction software packages were applied to analyze the KRP1 genome with respect to its genome plasticity. For GI detection the following programs were used: IslandViewer [98, 99], which incorporates the SIGI-HMM [23] and the IslandPath-DIMOB [24] software, and GIPSy (Genomic island prediction software) [26]. PHASTER (PHAge Search Tool – Enhanced Release) was used for identification and annotation of prophage sequences within the KRP1 genome [25, 30]. Spine and AGEnt were applied for prediction of the accessory genome in its entireness [100]. Results were imaged with BRIG (BLAST Ring Image Generator) [32]. This automated GI detection was complemented by manual curing of the precise starting and stopping position of each detected island via different blast comparisons. Additionally, the genome was manually mined for any of the 44 GI complexes previously described in *P. aeruginosa* [10-13].

The ACT (Artemis Comparison Tool) [50] was used for manual detection of regions of genomic plasticity (RGPs). It was also the visualization method of choice, for partial and whole genome comparisons of KRP1 with different *P. aeruginosa* strains.

Declarations

9.4 Competing interests
The authors declare that they have no competing interests
9.5 Funding
This work was supported by the Deutsche Forschungsgemeinschaft (DFG). The research grant AG156/1–1 from the DFG was awarded to MAR, CB was financed through this grant. The funding body did not play any roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.
9.6 Authors' Contributions
CB designed, executed and analyzed the experiments and drafted the manuscript, CR performed the sequence data clean up and mapping and published the finished genome. JB created the customized EDGAR project and ran the COG database comparison. KR isolated and provided the KRP1 strain. JK supervised the sequencing work. MAR conceived of the study, designed and interpreted the experiments and edited the manuscript. All authors revised the manuscript.
9.7 Acknowledgments
We thank Prof. Lars Blank for providing us access to his laboratory facilities and the helpful discussions at the iAMB in Aachen. We also want to express our gratitude to Dr. Tobias Busche and Anika Winkler from the CeBiTec in Bielefeld for generating the MinION library, performing the sequencing and in general for the excellent technical assistance.

Abbreviations

ANI: average nucleotide identity; GI: genomic island; ICE: integrative and conjugative element; LES: Liverpool Epidemic Strain; LESGI: Liverpool Epidemic Strain genomic island; MI: metabolic islands; ORF: open reading frame; PAGI: *Pseudomonas aeruginosa* genomic island; PAPI: *Pseudomonas aeruginosa* pathogenicity islands; PI: pathogenicity islands; RGP: regions of genomic plasticity; Rhs: rearrangement hot spot; RI: resistance islands; SI symbiotic islands.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The dataset (full genome data of *P. aeruginosa KRP1*) generated and analysed during the current study is available in the NCBI BioProject database repository, (http://www.ncbi.nlm.nih.gov/bioproject/) under accession number CP046069. It is part of the ElectricMicrobe100 Umbrella BioProject, which can be accessed via PRJNA417841.

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG). The research grant AG156/1-1 from the DFG was awarded to MAR, CB was financed through this grant. The funding body did not play any roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' Contributions

CB designed, executed and analyzed the experiments and drafted the manuscript, CR performed the sequence data clean up and mapping and published the finished genome. JB created the customized EDGAR project and ran the COG database comparison. KR isolated and provided the KRP1 strain. JK supervised the sequencing work. MAR conceived of the study, designed and interpreted the experiments and edited the manuscript. All authors revised the manuscript.

Acknowledgments

We thank Prof. Lars Blank for providing us access to his laboratory facilities and the helpful discussions at the iAMB in Aachen. We also want to express our gratitude to Dr. Tobias Busche and Anika Winkler from the CeBiTec in Bielefeld for generating the MinION library, performing the sequencing and in general for the excellent technical assistance.

References

1. Silby MW, Winstanley C, Godfrey SA, Levy SB, Jackson RW: ***Pseudomonas*** **genomes: diverse and adaptable**. *FEMS Microbiol Rev* 2011, **35**(4):652-680.
2. Hilker R, Munder A, Klockgether J, Losada PM, Chouvarine P, Cramer N, Davenport CF, Dethlefsen S, Fischer S, Peng H *et al*: **Interclonal gradient of virulence in the *Pseudomonas aeruginosa* pangenome**

from disease and environment. *Environ Microbiol* 2015, **17**(1):29-46.

3. de Bentzmann S, Plesiat P: **The *Pseudomonas aeruginosa* opportunistic pathogen and human infections**. *Environ Microbiol* 2011, **13**(7):1655-1665.
4. Lee DG, Urbach JM, Wu G, Liberati NT, Feinbaum RL, Miyata S, Diggins LT, He J, Saucier M, Deziel E *et al*: **Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial**. *Genome Biol* 2006, **7**(10):R90.
5. He J, Baldini RL, Déziel E, Saucier M, Zhang Q, Liberati NT, Lee D, Urbach J, Goodman HM, Rahme LG: **The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes**. *Proc Natl Acad Sci U S A* 2004, **101**(8):2530-2535.
6. Bennett PM: **Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangement**. *Methods Mol Biol* 2004, **266**:71-113.
7. Kung VL, Ozer EA, Hauser AR: **The Accessory Genome of *Pseudomonas aeruginosa***. *Microbiol Mol Biol Rev* 2010, **74**(4):621-641.
8. Tümmler B: **Clonal Variations in *Pseudomonas aeruginosa***. In: *Pseudomonas: Volume 4 Molecular Biology of Emerging Issues.* Edited by Ramos J-L, Levesque RC. Boston, MA: Springer US; 2006: 35-68.
9. Hooper SD, Berg OG: **On the nature of gene innovation: duplication patterns in microbial genomes**. *Mol Biol Evol* 2003, **20**(6):945-954.
10. Klockgether J, Cramer N, Wiehlmann L, Davenport CF, Tummler B: ***Pseudomonas aeruginosa* Genomic Structure and Diversity**. *Front Microbiol* 2011, **2**:150.
11. Silveira MC, Albano RM, Asensi MD, Carvalho-Assef APDA: **Description of genomic islands associated to the multidrug-resistant *Pseudomonas aeruginosa* clone ST277**. *Infect Genet Evol* 2016, **42**:60-65.
12. Hong JS, Yoon EJ, Lee H, Jeong SH, Lee K: **Clonal Dissemination of *Pseudomonas aeruginosa* Sequence Type 235 Isolates Carrying blaIMP-6 and Emergence of blaGES-24 and blaIMP-10 on Novel Genomic Islands PAGI-15 and -16 in South Korea**. *Antimicrob Agents Chemother* 2016, **60**(12):7216-7223.
13. Jani M, Mathee K, Azad RK: **Identification of Novel Genomic Islands in Liverpool Epidemic Strain of *Pseudomonas aeruginosa* Using Segmentation and Clustering**. *Front Microbiol* 2016, **7**:1210.
14. Battle SE, Meyer F, Rello J, Kung VL, Hauser AR: **Hybrid Pathogenicity Island PAGI-5 Contributes to the Highly Virulent Phenotype of a *Pseudomonas aeruginosa* Isolate in Mammals**. *J Bacteriol* 2008, **190**(21):7130-7140.
15. Kung VL, Khare S, Stehlik C, Bacon EM, Hughes AJ, Hauser AR: **An *rhs* gene of *Pseudomonas aeruginosa* encodes a virulence protein that activates the inflammasome**. *Proc Natl Acad Sci U S A* 2012, **109**(4):1275-1280.
16. Winstanley C, Langille MG, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C, Sanschagrin F, Thomson NR, Winsor GL, Quail MA, Lennard N *et al*: **Newly introduced genomic prophage islands are critical**

determinants of *in vivo* competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. *Genome Res* 2009, **19**(1):12-23.

17. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FSL, Hufnagle WO, Kowalik DJ, Lagrou M *et al*: **Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen**. *Nature* 2000, **406**(6799):959-964.
18. Déraspe M, Alexander DC, Xiong J, Ma JH, Low DE, Jamieson FB, Roy PH: **Genomic analysis of *Pseudomonas aeruginosa* PA96, the host of carbapenem resistance plasmid pOZ176**. *FEMS Microbiol Lett* 2014, **356**(2):212-216.
19. Blom J, Albaum SP, Doppmeier D, Pühler A, Vorhölter F-J, Zakrzewski M, Goesmann A: **EDGAR: A software framework for the comparative analysis of prokaryotic genomes**. *BMC Bioinf* 2009, **10**(1):154.
20. Blom J, Kreis J, Spänig S, Juhre T, Bertelli C, Ernst C, Goesmann A: **EDGAR 2.0: an enhanced software platform for comparative gene content analyses**. *Nucleic Acids Res* 2016, **44**(W1):W22-W28.
21. Ali F, Hu H, Xu P, Tang H: **Complete Genome Sequence of *Pseudomonas aeruginosa* FA-HZ1, an Efficient Dibenzofuran-Degrading Bacterium**. *Genome Announc* 2017, **5**(7):e01634-01616.
22. Yan J, Deforet M, Boyle KE, Rahman R, Liang R, Okegbe C, Dietrich LEP, Qiu W, Xavier JB: **Bow-tie signaling in c-di-GMP: Machine learning in a simple biochemical network**. *PLoS Comput Biol* 2017, **13**(8):e1005677.
23. Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, Surovcik K, Meinicke P, Merkl R: **Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models**. *BMC Bioinf* 2006, **7**(1):142.
24. Hsiao W, Wan I, Jones SJ, Brinkman FSL: **IslandPath: aiding detection of genomic islands in prokaryotes**. *Bioinformatics* 2003, **19**(3):418-420.
25. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS: **PHASTER: a better, faster version of the PHAST phage search tool**. *Nucleic Acids Res* 2016, **44**(Web Server issue):W16-W21.
26. Soares SC, Geyik H, Ramos RTJ, de Sá PHCG, Barbosa EGV, Baumbach J, Figueiredo HCP, Miyoshi A, Tauch A, Silva A *et al*: **GIPSy: Genomic island prediction software**. *J Biotechnol* 2016, **232**:2-11.
27. Mathee K, Narasimhan G, Valdes C, Qiu X, Matewish JM, Koehrsen M, Rokas A, Yandava CN, Engels R, Zeng E *et al*: **Dynamics of *Pseudomonas aeruginosa* genome evolution**. *Proc Natl Acad Sci U S A* 2008, **105**(8):3100-3105.
28. Smith EE, Sims EH, Spencer DH, Kaul R, Olson MV: **Evidence for Diversifying Selection at the Pyoverdine Locus of *Pseudomonas aeruginosa***. *J Bacteriol* 2005, **187**(6):2138-2147.
29. Subedi D, Kohli GS, Vijay AK, Willcox M, Rice SA: **Accessory genome of the multi-drug resistant ocular isolate of *Pseudomonas aeruginosa* PA34**. *PLoS One* 2019, **14**(4):e0215038.
30. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS: **PHAST: A Fast Phage Search Tool**. *Nucleic Acids Res* 2011, **39**(Web Server issue):W347-W352.

31. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL: **Prophages mediate defense against phage infection through diverse mechanisms**. *The ISME Journal* 2016, **10**(12):2854-2866.
32. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA: **BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons**. *BMC Genomics* 2011, **12**(1):402.
33. Roy PH, Tetu SG, Larouche A, Elbourne L, Tremblay S, Ren Q, Dodson R, Harkins D, Shay R, Watkins K *et al*: **Complete genome sequence of the multiresistant taxonomic outlier *Pseudomonas aeruginosa* PA7**. *PLoS One* 2010, **5**(1):e8842.
34. Battle SE, Rello J, Hauser AR: **Genomic Islands of *Pseudomonas aeruginosa***. *FEMS Microbiol Lett* 2009, **290**(1):70-78.
35. Boukerb AM, Decor A, Ribun S, Tabaroni R, Rousset A, Commin L, Buff S, Doléans-Jordheim A, Vidal S, Varrot A *et al*: **Genomic Rearrangements and Functional Diversification of *lecA* and *lecB* Lectin-Coding Regions Impacting the Efficacy of Glycomimetics Directed against *Pseudomonas aeruginosa***. *Front Microbiol* 2016, **7**:811.
36. Liu PV, Wang S: **Three new major somatic antigens of *Pseudomonas aeruginosa***. *J Clin Microbiol* 1990, **28**(5):922-925.
37. Burrows LL, Charter DF, Lam JS: **Molecular characterization of the *Pseudomonas aeruginosa* serotype O5 (PAO1) B-band lipopolysaccharide gene cluster**. *Mol Microbiol* 1996, **22**(3):481-495.
38. Lam JS, Taylor VL, Islam ST, Hao Y, Kocíncová D: **Genetic and Functional Diversity of *Pseudomonas aeruginosa* Lipopolysaccharide**. *Front Microbiol* 2011, **2**:118-118.
39. Meyer JM, Stintzi A, De Vos D, Cornelis P, Tappe R, Taraz K, Budzikiewicz H: **Use of siderophores to type pseudomonads: the three *Pseudomonas aeruginosa* pyoverdine systems**. *Microbiology* 1997, **143 (Pt 1)**:35-43.
40. Kus JV, Tullis E, Cvitkovitch DG, Burrows LL: **Significant differences in type IV pilin allele distribution among *Pseudomonas aeruginosa* isolates from cystic fibrosis (CF) versus non-CF patients**. *Microbiology* 2004, **150**(5):1315-1326.
41. Giltner CL, Rana N, Lunardo MN, Hussain AQ, Burrows LL: **Evolutionary and functional diversity of the Pseudomonas type IVa pilin island**. 2011, **13**(1):250-264.
42. Arora SK, Bangera M, Lory S, Ramphal R: **A genomic island in *Pseudomonas aeruginosa* carries the determinants of flagellin glycosylation**. *Proc Natl Acad Sci U S A* 2001, **98**(16):9342-9347.
43. Verma A, Schirm M, Arora SK, Thibault P, Logan SM, Ramphal R: **Glycosylation of b-Type Flagellin of *Pseudomonas aeruginosa:* Structural and Genetic Basis**. *J Bacteriol* 2006, **188**(12):4395-4403.
44. Varga JJ, Barbier M, Mulet X, Bielecki P, Bartell JA, Owings JP, Martinez-Ramos I, Hittle LE, Davis MR, Damron FH *et al*: **Genotypic and phenotypic analyses of a *Pseudomonas aeruginosa* chronic bronchiectasis isolate reveal differences from cystic fibrosis and laboratory strains**. *BMC Genomics* 2015, **16**(1):883.
45. Nakayama K, Kanaya S, Ohnishi M, Terawaki Y, Hayashi T: **The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa:* implications for phage evolution**

and horizontal gene transfer via bacteriophages. *Mol Microbiol* 1999, **31**(2):399-419.

46. Cazares A, Mendoza-Hernández G, Guarneros G: **Core and accessory genome architecture in a group of *Pseudomonas aeruginosa* Mu-like phages**. *BMC Genomics* 2014, **15**(1):1146.
47. Larbig KD, Christmann A, Johann A, Klockgether J, Hartsch T, Merkl R, Wiehlmann L, Fritz H-J, Tümmler B: **Gene Islands Integrated into tRNA(Gly) Genes Confer Genome Diversity on a *Pseudomonas aeruginosa* Clone**. *J Bacteriol* 2002, **184**(23):6665-6680.
48. Kiewitz C, Larbig K, Klockgether J, Weinel C, Tümmler B: **Monitoring genome evolution *ex vivo*: reversible chromosomal integration of a 106 kb plasmid at two tRNALys gene loci in sequential *Pseudomonas aeruginosa* airway isolates**. *Microbiology* 2000, **146**(10):2365-2373.
49. Mohd-Zain Z, Turner SL, Cerdeño-Tárraga AM, Lilley AK, Inzana TJ, Duncan AJ, Harding RM, Hood DW, Peto TE, Crook DW: **Transferable Antibiotic Resistance Elements in *Haemophilus influenzae* Share a Common Evolutionary Origin with a Diverse Family of Syntenic Genomic Islands**. *J Bacteriol* 2004, **186**(23):8114-8122.
50. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J: **ACT: the Artemis Comparison Tool**. *Bioinformatics* 2005, **21**(16):3422-3423.
51. Clewell DB: **Bacterial Conjugation**, 1 edn: Springer US; 1993.
52. Cheng K, Smyth RL, Govan JR, Doherty C, Winstanley C, Denning N, Heaf DP, van Saene H, Hart CA: **Spread of beta-lactam-resistant *Pseudomonas aeruginosa* in a cystic fibrosis clinic**. *Lancet* 1996, **348**(9028):639-642.
53. Hachani A, Allsopp LP, Oduko Y, Filloux A: **The VgrG Proteins Are "à la Carte" Delivery Systems for Bacterial Type VI Effectors**. *J Biol Chem* 2014, **289**(25):17872-17884.
54. Jones C, Hachani A, Manoli E, Filloux A: **An *rhs* Gene Linked to the Second Type VI Secretion Cluster Is a Feature of the *Pseudomonas aeruginosa* Strain PA14**. *J Bacteriol* 2014, **196**(4):800-810.
55. Whitney JC, Beck CM, Goo YA, Russell AB, Harding BN, Leon JAD, Cunningham DA, Tran BQ, Low DA, Goodlett DR *et al*: **Genetically distinct pathways guide effector export through the type VI secretion system**. *Mol Microbiol* 2014, **92**(3):529-542.
56. Jamet A, Nassif X: **New Players in the Toxin Field: Polymorphic Toxin Systems in Bacteria**. *mBio* 2015, **6**(3).
57. Seemann T: **Prokka: rapid prokaryotic genome annotation**. *Bioinformatics* 2014, **30**(14):2068-2069.
58. Hauser AR, Cobb E, Bodi M, Mariscal D, Valles J, Engel JN, Rello J: **Type III protein secretion is associated with poor clinical outcomes in patients with ventilator-associated pneumonia caused by *Pseudomonas aeruginosa***. *Crit Care Med* 2002, **30**(3):521-528.
59. Schulert GS, Feltman H, Rabin SDP, Martin CG, Battle SE, Rello J, Hauser AR: **Secretion of the Toxin ExoU Is a Marker for Highly Virulent *Pseudomonas aeruginosa* Isolates Obtained from Patients with Hospital-Acquired Pneumonia**. *J Infect Dis* 2003, **188**(11):1695-1706.
60. Pier GB: ***Pseudomonas aeruginosa* lipopolysaccharide: a major virulence factor, initiator of inflammation and target for effective immunity**. *Int J Med Microbiol* 2007, **297**(5):277-295.

61. Juan C, Peña C, Oliver A: **Host and Pathogen Biomarkers for Severe *Pseudomonas aeruginosa* Infections**. *J Infect Dis* 2017, **215**(suppl_1):S44-S51.
62. Rabaey K, Boon N, Siciliano SD, Verhaege M, Verstraete W: **Biofuel cells select for microbial consortia that self-mediate electron transfer**. *Appl Environ Microbiol* 2004, **70**(9):5373-5382.
63. Kiewitz C, Tümmler B: **Sequence Diversity of *Pseudomonas aeruginosa*: Impact on Population Structure and Genome Evolution**. *J Bacteriol* 2000, **182**(11):3125-3135.
64. Spencer DH, Kas A, Smith EE, Raymond CK, Sims EH, Hastings M, Burns JL, Kaul R, Olson MV: **Whole-genome sequence variation among multiple isolates of *Pseudomonas aeruginosa***. *J Bacteriol* 2003, **185**(4):1316-1325.
65. Cramer N, Klockgether J, Wrasman K, Schmidt M, Davenport CF, Tummler B: **Microevolution of the major common *Pseudomonas aeruginosa* clones C and PA14 in cystic fibrosis lungs**. *Environ Microbiol* 2011, **13**(7):1690-1704.
66. Valot B, Guyeux C, Rolland JY, Mazouzi K, Bertrand X, Hocquet D: **What It Takes to Be a *Pseudomonas aeruginosa*? The Core Genome of the Opportunistic Pathogen Updated**. *PLoS One* 2015, **10**(5):e0126468.
67. Baysse C, Meyer J-M, Plesiat P, Geoffroy V, Michel-Briand Y, Cornelis P: **Uptake of Pyocin S3 Occurs through the Outer Membrane Ferripyoverdine Type II Receptor of *Pseudomonas aeruginosa***. *J Bacteriol* 1999, **181**(12):3849-3851.
68. Hahn HP: **The type-4 pilus is the major virulence-associated adhesin of *Pseudomonas aeruginosa*–a review**. *Gene* 1997, **192**(1):99-108.
69. Sriramulu DD, Lunsdorf H, Lam JS, Romling U: **Microcolony formation: a novel biofilm model of *Pseudomonas aeruginosa* for the cystic fibrosis lung**. *J Med Microbiol* 2005, **54**(Pt 7):667-676.
70. Bradley DE, Pitt TL: **Pilus-dependence of four *Pseudomonas aeruginosa* bacteriophages with non-contractile tails**. *J Gen Virol* 1974, **24**(1):1-15.
71. Roncero C, Darzins A, Casadaban M: ***Pseudomonas aeruginosa* transposable bacteriophages D3112 and B3 require pili and surface growth for adsorption**. *J Bacteriol* 1990, **172**(4):1899-1904.
72. Harvey H, Bondy-Denomy J, Marquis H, Sztanko KM, Davidson AR, Burrows LL: ***Pseudomonas aeruginosa* defends against phages through type IV pilus glycosylation**. *Nat Microbiol* 2018, **3**(1):47-52.
73. Schirm M, Arora SK, Verma A, Vinogradov E, Thibault P, Ramphal R, Logan SM: **Structural and Genetic Characterization of Glycosylation of Type a Flagellin in *Pseudomonas aeruginosa***. *J Bacteriol* 2004, **186**(9):2523-2531.
74. Arora SK, Neely AN, Blair B, Lory S, Ramphal R: **Role of Motility and Flagellin Glycosylation in the Pathogenesis of *Pseudomonas aeruginosa* Burn Wound Infections**. *Infect Immun* 2005, **73**(7):4395-4398.
75. Kuang Z, Hao Y, Hwang S, Zhang S, Kim E, Akinbi HT, Schurr MJ, Irvin RT, Hassett DJ, Lau GW: **The *Pseudomonas aeruginosa* flagellum confers resistance to pulmonary surfactant protein-A by**

impacting the production of exoproteases through quorum-sensing. *Mol Microbiol* 2011, **79**(5):1220-1235.

76. Bultreys A: **Siderotyping, a Tool to Characterize, Classify and Identify Fluorescent Pseudomonads**. In: *Microbial Siderophores.* Edited by Varma A, Chincholkar SB. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007: 67-89.
77. Meyer JM, Neely A, Stintzi A, Georges C, Holder IA: **Pyoverdin is essential for virulence of *Pseudomonas aeruginosa***. *Infect Immun* 1996, **64**(2):518-523.
78. Cornelis P, Dingemans J: ***Pseudomonas aeruginosa* adapts its iron uptake strategies in function of the type of infections**. *Front Cell Infect Microbiol* 2013, **3**:75.
79. Augustin DK, Song Y, Baek MS, Sawa Y, Singh G, Taylor B, Rubio-Mills A, Flanagan JL, Wiener-Kronish JP, Lynch SV: **Presence or Absence of Lipopolysaccharide O Antigens Affects Type III Secretion by *Pseudomonas aeruginosa***. *J Bacteriol* 2007, **189**(6):2203-2209.
80. Bosire EM, Blank LM, Rosenbaum MA: **Strain and substrate dependent redox mediator and electricity production by *Pseudomonas aeruginosa***. *Appl Environ Microbiol* 2016:AEM. 01342-01316.
81. Price-Whelan A, Dietrich LEP, Newman DK: **Rethinking 'secondary' metabolism: physiological roles for phenazine antibiotics**. *Nat Chem Biol* 2006, **2**(2):71-78.
82. Caldwell CC, Chen Y, Goetzmann HS, Hao Y, Borchers MT, Hassett DJ, Young LR, Mavrodi D, Thomashow L, Lau GW: ***Pseudomonas aeruginosa* Exotoxin Pyocyanin Causes Cystic Fibrosis Airway Pathogenesis**. *Am J Pathol* 2009, **175**(6):2473-2488.
83. Wolfgang MC, Kulasekara BR, Liang X, Boyd D, Wu K, Yang Q, Miyada CG, Lory S: **Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa***. *Proc Natl Acad Sci U S A* 2003, **100**(14):8484-8489.
84. Panagea S, Winstanley C, Walshaw MJ, Ledson MJ, Hart CA: **Environmental contamination with an epidemic strain of *Pseudomonas aeruginosa* in a Liverpool cystic fibrosis centre, and study of its survival on dry surfaces**. *J Hosp Infect* 2005, **59**(2):102-107.
85. Al-Aloul M, Crawley J, Winstanley C, Hart CA, Ledson MJ, Walshaw MJ: **Increased morbidity associated with chronic infection by an epidemic *Pseudomonas aeruginosa* strain in CF patients**. *Thorax* 2004, **59**(4):334-336.
86. Salunkhe P, Smart CH, Morgan JA, Panagea S, Walshaw MJ, Hart CA, Geffers R, Tummler B, Winstanley C: **A cystic fibrosis epidemic strain of *Pseudomonas aeruginosa* displays enhanced virulence and antimicrobial resistance**. *J Bacteriol* 2005, **187**(14):4908-4920.
87. Fothergill JL, Panagea S, Hart CA, Walshaw MJ, Pitt TL, Winstanley C: **Widespread pyocyanin over-production among isolates of a cystic fibrosis epidemic strain**. *BMC Microbiol* 2007, **7**:45.
88. McCallum SJ, Corkill J, Gallagher M, Ledson MJ, Hart CA, Walshaw MJ: **Superinfection with a transmissible strain of *Pseudomonas aeruginosa* in adults with cystic fibrosis chronically colonised by *P. aeruginosa***. *Lancet* 2001, **358**(9281):558-560.
89. McCallum S, Gallagher M, Corkill J, Hart C, Ledson M, Walshaw M: **Spread of an epidemic *Pseudomonas aeruginosa* strain from a patient with cystic fibrosis (CF) to non-CF relatives**. *Thorax*

2002, **57**(6):559-560.

90. Harrison EM, Carter MEK, Luck S, Ou H-Y, He X, Deng Z, O'Callaghan C, Kadioglu A, Rajakumar K: **Pathogenicity Islands PAPI-1 and PAPI-2 Contribute Individually and Synergistically to the Virulence of *Pseudomonas aeruginosa* Strain PA14**. *Infect Immun* 2010, **78**(4):1437-1446.
91. Altenbuchner J, Cullum J: **DNA amplification and an unstable arginine gene in *Streptomyces lividans* 66**. *Mol Gen Genet* 1984, **195**(1):134-138.
92. **https: //github.com/Albacore/albacore**
93. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM: **Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation**. *Genome Res* 2017, **27**(5):722-736.
94. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK *et al*: **Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement**. *PLoS One* 2014, **9**(11):e112963.
95. Gordon D, Green P: **Consed: a graphical editor for next-generation sequencing**. *Bioinformatics* 2013, **29**(22):2936-2937.
96. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: **Artemis: sequence visualization and annotation**. *Bioinformatics (Oxford, England)* 2000, **16**(10):944-945.
97. Tatusov RL, Galperin MY, Natale DA, Koonin EV: **The COG database: a tool for genome-scale analysis of protein functions and evolution**. *Nucleic Acids Res* 2000, **28**(1):33-36.
98. Langille MGI, Brinkman FSL: **IslandViewer: an integrated interface for computational identification and visualization of genomic islands**. *Bioinformatics* 2009, **25**(5):664-665.
99. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, Brinkman FSL: **IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets**. *Nucleic Acids Res* 2017, **45**(W1):W30-w35.
100. Ozer EA, Allen JP, Hauser AR: **Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt**. *BMC Genomics* 2014, **15**(1):737.

Supplementary Information

Table S1: Calculated average nucleotide identity (ANI) values for 104 *P. aeruginosa* strains and 8 other *Pseudomonas* strains vs. *P. aeruginosa* KRP1. Shown are ANI values of different *Pseudomonas* strains versus the *P. aeruginosa* KRP1 strain. Values are given in %. The higher the respective number, the greater is the nucleotide identity between the respective genomes.

Table S2: Clusters of Orthologous Groups (COG) classification of singleton hits of *P. aeruginosa* KRP1. Singletons of *P. aeruginosa* KRP1 are denoted with their respective locus tag. The COG categories and the exact COG annotations are shown, for all singletons of KRP1 with respect to *P. aeruginosa* PAO1, PA14, LESB58, FA-HZ1 and W45909, which could be identified within the COG database. The percent identity of the respective singletons with the deposited database is also given.

Table S3: Absolute number of open reading frames (ORFs) belonging to different Clusters of Orthologous Groups (COG) database categories of areas I to VI. Areas correspond to groups of genes, which are singletons to KRP1 or shared by KRP1 and up to five other *P. aeruginosa* strains; see Figure 4**.** Official COG categories are denoted and for each category the absolute number of genes for each area are given.

Table S4: PHASTER prediction for putative genomic islands (GIs) of KRP1 characterized as prophages and phage-like elements. The GI numbers given correspond to the GIs introduced in Table 2. For each prophages and phage-like elements the phage with the highest similarity is given, and the number of homologues ORFs, found in the PHASTER database is stated in parentheses. Additionally, the software provides a prediction if the phage is intact, questionable or incomplete, depending on the hit percent, also shown within the table.

Figure S1: Phylogenetic tree of 105 fully sequenced *P. aeruginosa* strains. *P. aeruginosa* KRP1 (red), the two closest relatives *P. aeruginosa* FA-HZ1 and W45909 (blue), the model strain PAO1, the common research strain PA14 and the potent virulent strain LESB58 (green) and the strain ATCC 27853 (pink), which showed the highest 16S rRNA similarity in the original publication from [62], are highlighted. The tree was calculated using the EDGAR platform [19, 20] out of a core of 2,567 genes per genome, which transcribe to 780,580 amino acid residues per genome.

Figures

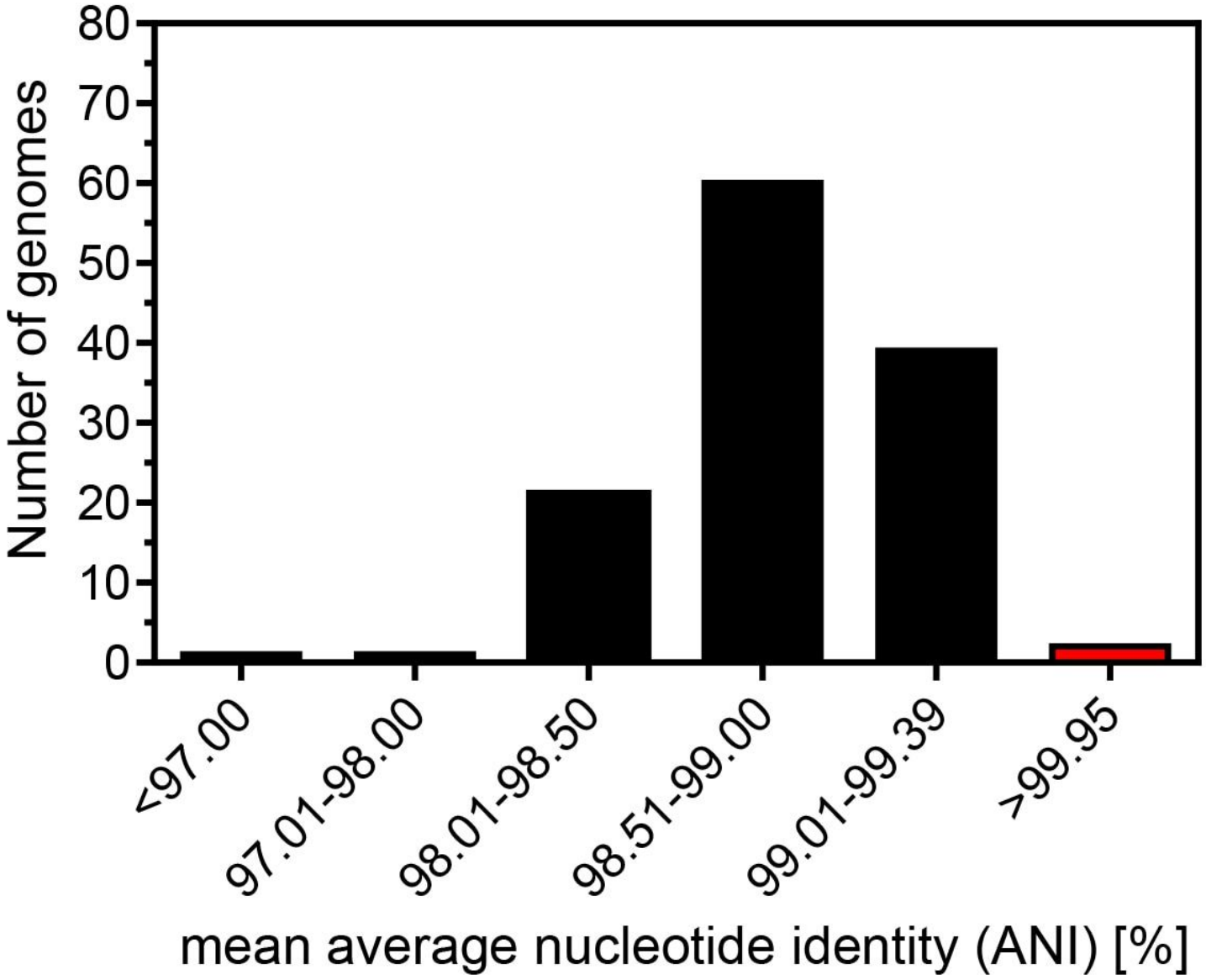

Figure 1

Average nucleotide identity (ANI) percentage of 105 P. aeruginosa genomes vs. the genome of P. aeruginosa KRP1. Values were generated using the EDGAR platform [19, 20]. The highest similarities are found between P. aeruginosa KRP1 and P. aeruginosa W45909 (ANI = 99.96%) as well as P. aeruginosa FA-HZ1 (ANI = 99.98%) (red column). For exact values see Supplement information Table S1.

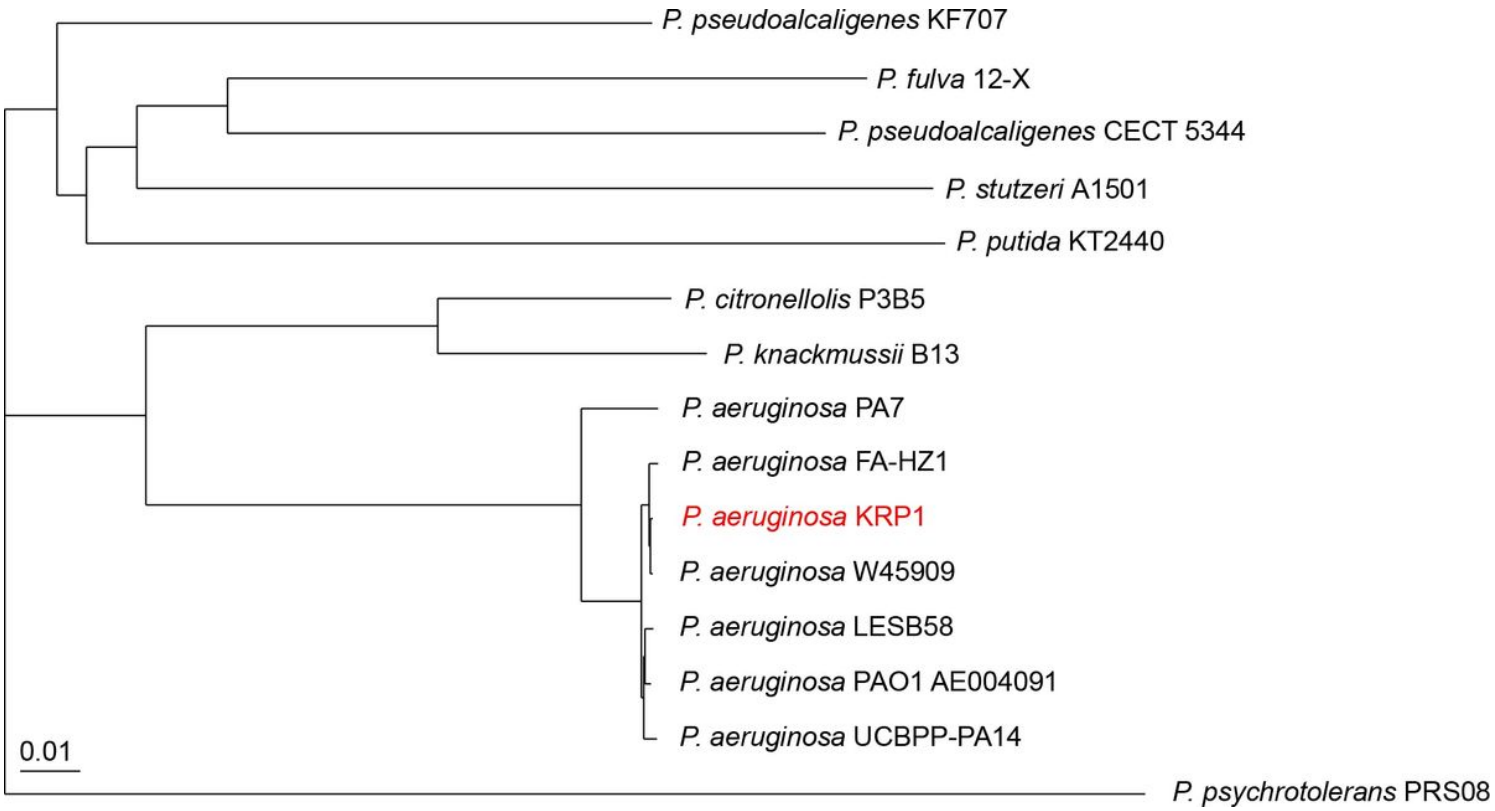

Figure 2

Phylogenetic tree of six fully sequenced P. aeruginosa strains and eight other Pseudomonas species. The tree was calculated using the EDGAR platform [19, 20] out of a core of 1,537 genes per genome comprised of 532,537 amino acid-residues per genome.

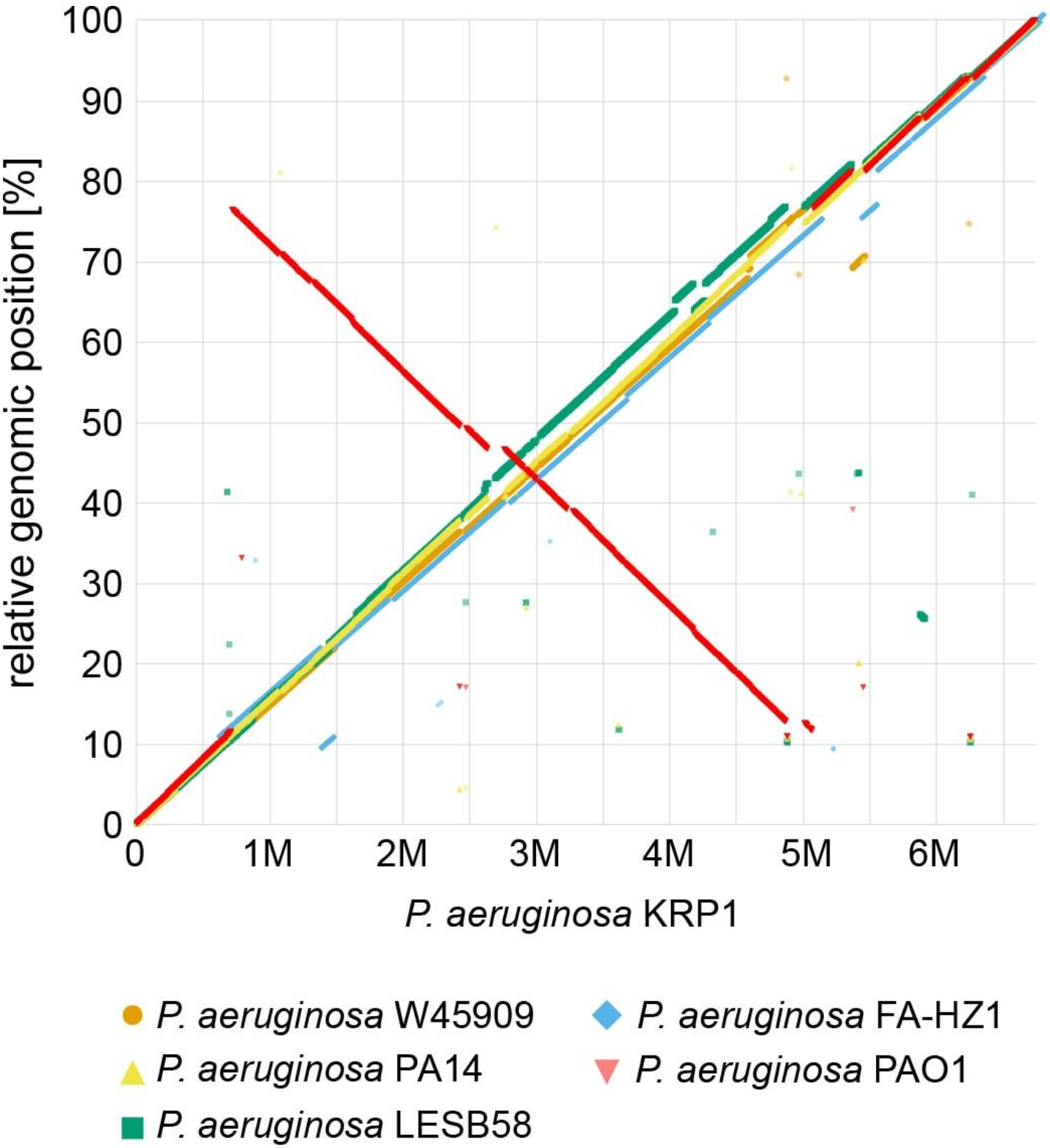

Figure 3

Synteny plot of the P. aeruginosa KRP1 strain and five other P. aeruginosa strains. Each dot represents a gene of KRP1 and its corresponding homologue in the respective comparative strain. X-axis shows the position within the chromosome of KRP1 and y-axis shows the relative position within the compared genome. Analysis was performed with the EDGAR platform [19, 20]

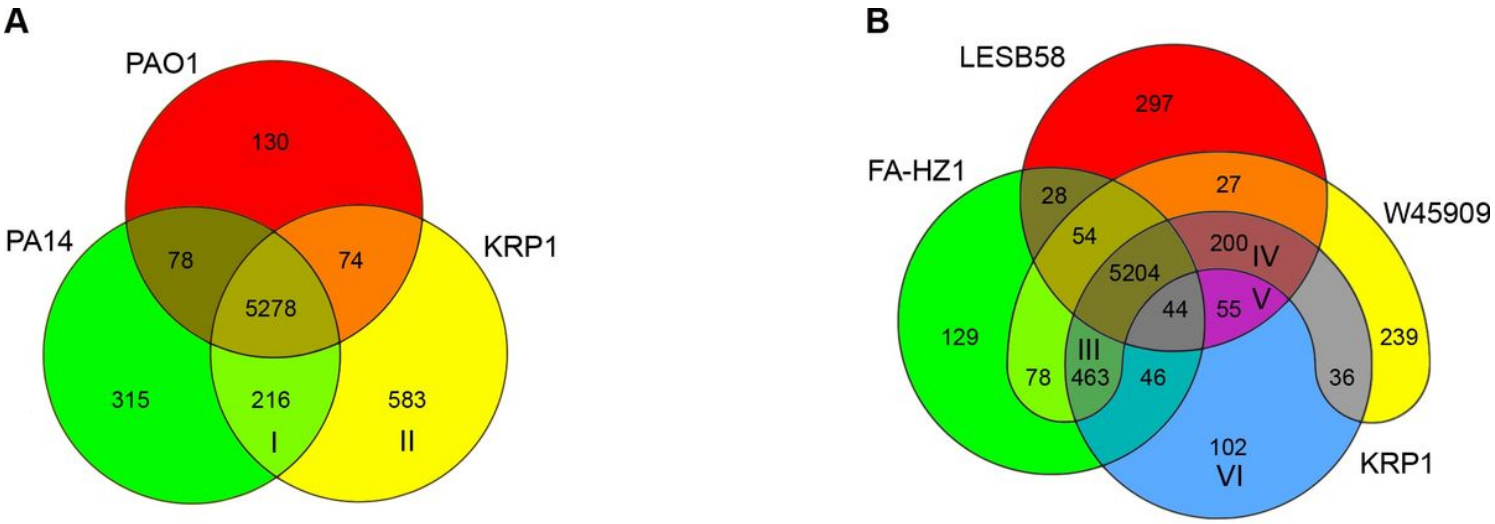

Figure 4

Venn diagrams showing the number of genes shared as orthologues in all possible logical combinations between different strains of P. aeruginosa. A: PAO1 [red], PA14 [green], KRP1 [blue]; B: LESB58 [red], FA-HZ1 [green], KRP1 [blue], W45909 [yellow]. For further information regarding individual areas marked with Roman numbers see text. Analysis was performed with the EDGAR platform [19, 20]

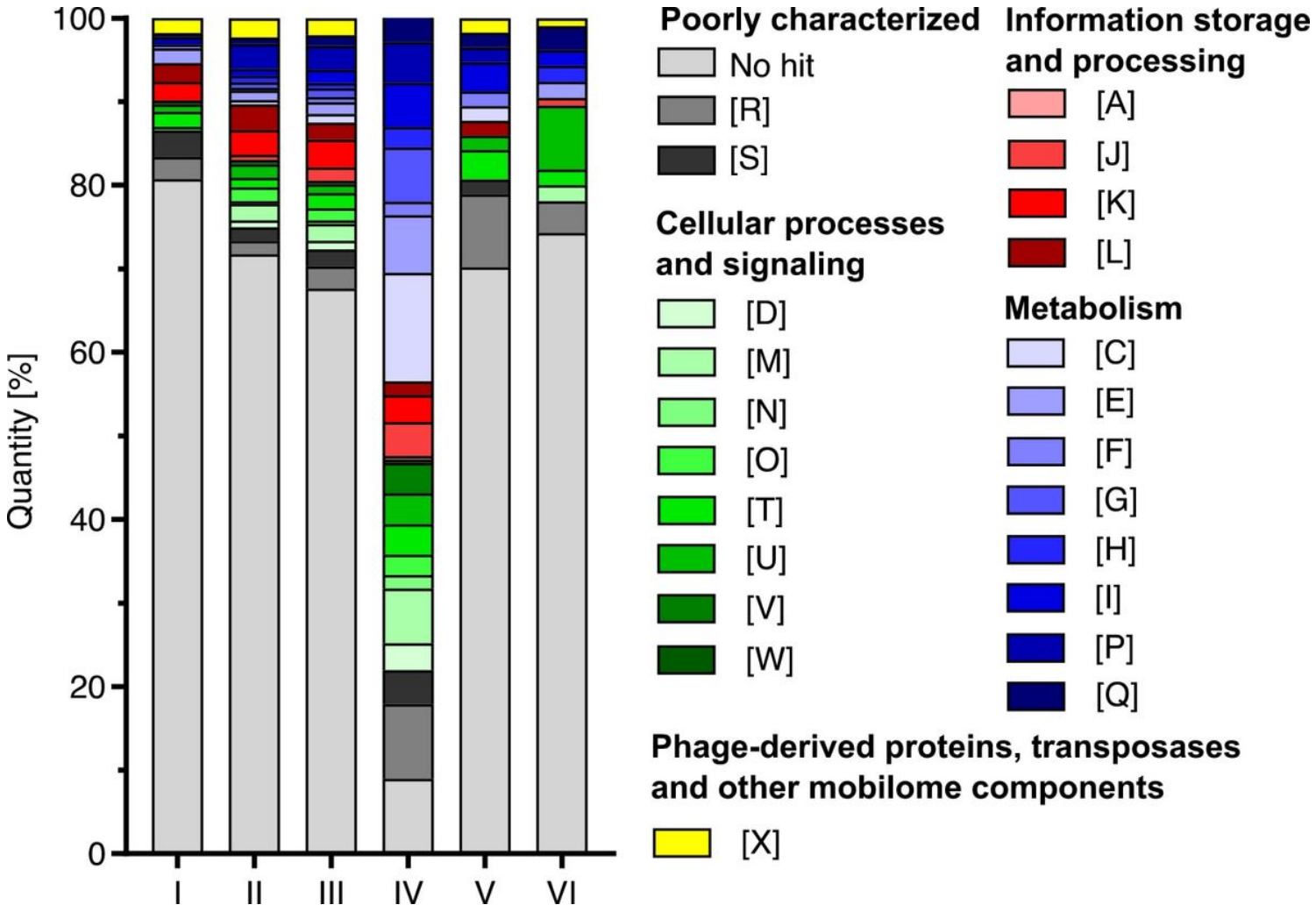

Figure 5

ORFs of areas I to VI (groups of genes, which are singletons to KRP1 or shared by KRP1 and up to five other P. aeruginosa strains; see Figure 4) classified by Clusters of Orthologous Groups (COGs) database. Category designations are as follows: [R] - General function prediction only; [S] - Function unknown; [D] - Cell cycle control, cell division, chromosome partitioning; [M] - Cell wall/membrane/envelope biogenesis; [N] - Cell motility; [O] - Post-translational modification, protein turnover, and chaperones; [T] - Signal transduction mechanisms; [U] - Intracellular trafficking, secretion, and vesicular transport; [V] - Defense mechanisms; [W] - Extracellular structures; [A] - RNA processing and modification; [J] - Translation, ribosomal structure and biogenesis; [K] – Transcription; [L] - Replication, recombination and repair; [C] - Energy production and conversion; [E] - Amino acid transport and metabolism; [F] - Nucleotide transport and metabolism; [G] - Carbohydrate transport and metabolism; [H] - Coenzyme transport and metabolism; [I] - Lipid transport and metabolism; [P] - Inorganic ion transport and metabolism; [Q] - Secondary metabolites biosynthesis, transport, and catabolism; [X] - Phage-derived proteins, transposases and other mobilome components.

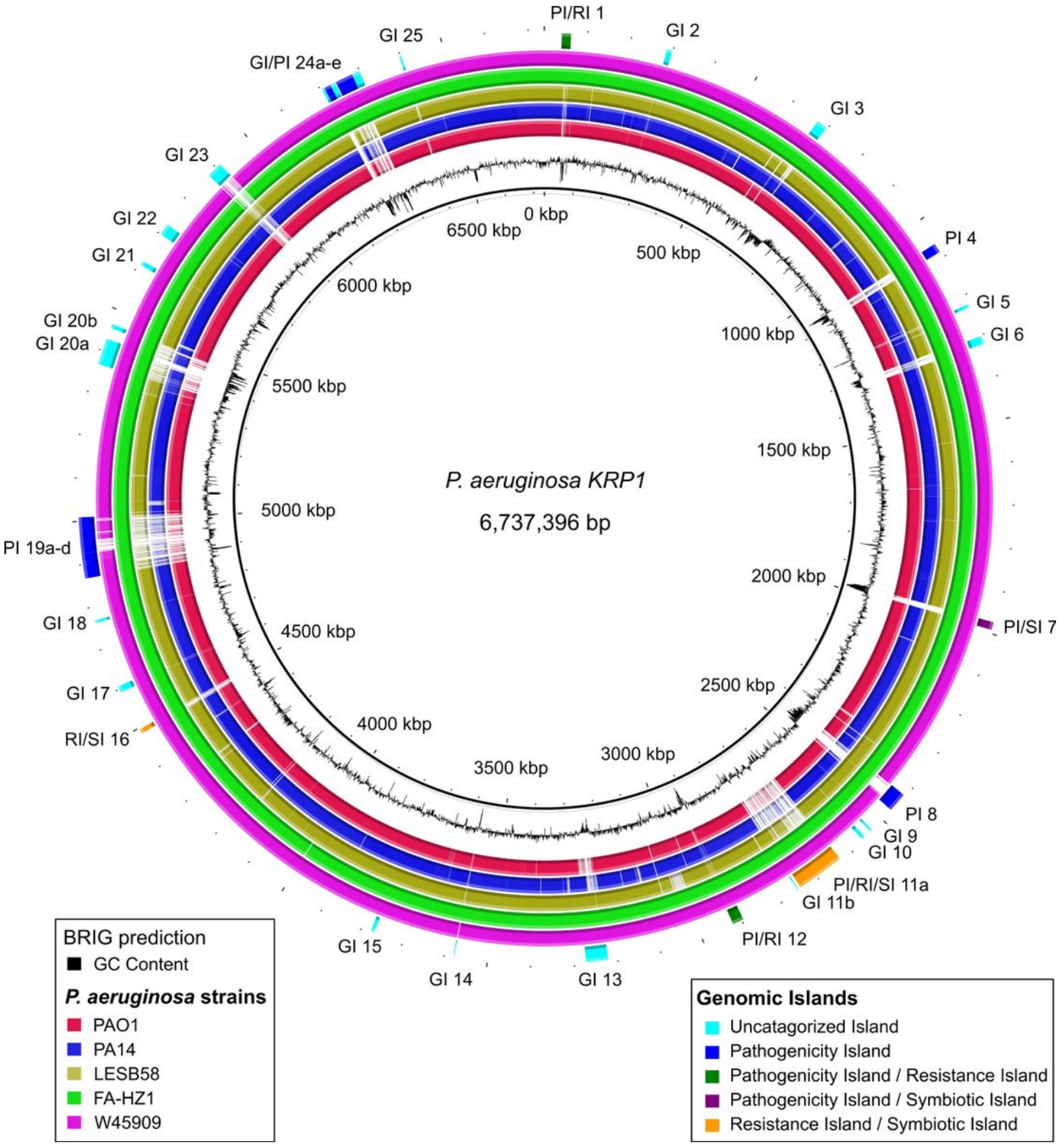

Figure 6

Visualization of genome plasticity in the P. aeruginosa KRP1 genome detected with different prediction programs. KRP1 main chromosome in comparison to selected P. aeruginosa genomes. Starting from the innermost circle going outwards: major (500kbp) and minor tick (100kbp) measurements of the KRP1 genome; G+C content (black traces); BLAST comparisons of PAO1 genome against the KRP1 genome (red ring); BLAST comparisons of PA14 genome against the KRP1 genome (blue ring); BLAST

comparisons of LESB58 genome against the KRP1 genome (ocher ring) BLAST comparisons of FA-HZ1 genome against the KRP1 genome (green ring); BLAST comparisons of W45909 genome against the KRP1 genome (magenta ring); combined genome plasticity prediction of SIGI-HMM [23], IslandPath-DIMOB [24], PHASTER [25] and GIPSy [26], when comparing KRP1 to PA14 as a reference (aqua segments: uncategorized genomic islands [GIs]; blue segments: pathogenicity islands [PIs]; green segments: pathogenicity / resistance islands [PI/RIs]; purple segments: pathogenicity / symbiotic islands [PI/SIs]; orange segments: resistance / symbiotic islands [RI/SIs]). Whole genome BLAST comparison and image generation was performed with BRIG [32].

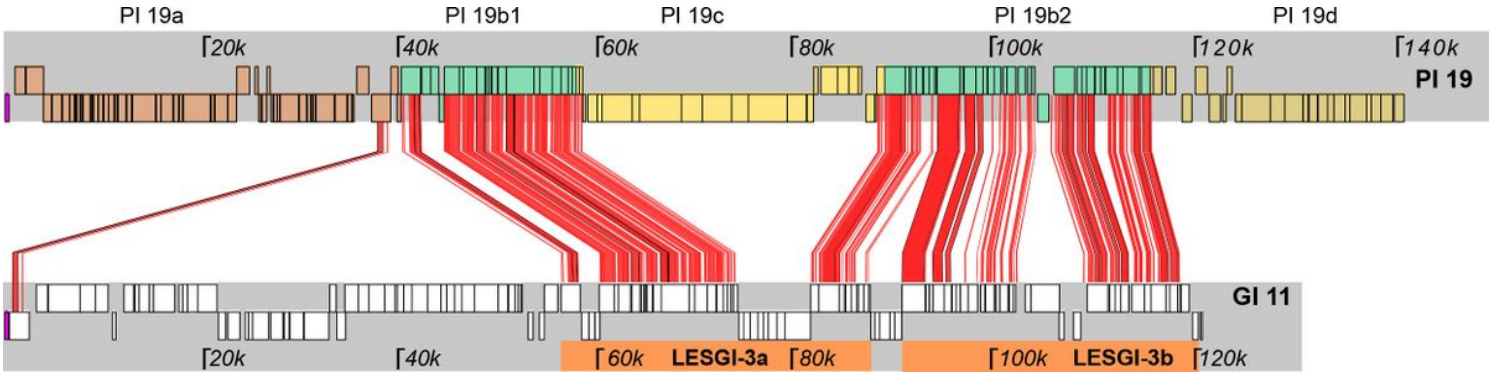

Figure 7

ACT BLAST comparison [50] of Pathogenicity island (PI) 19 of P. aeruginosa KRP1 with genomic island (GI) 11 of P. aeruginosa KRP1. Boxes highlight ORFs and the color code in PI 19 differentiates the different evolutionary sub-sections of PI 19. Integration site tRNAs are marked in pink. LESGI-3 a & b refer to a GI of P. aeruginosa strain LESB58 [16]. Red lines connecting the two islands mark blast matches between the different sequences.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.docx
- TableS3.docx
- TableS1.docx
- TableS4.docx
- TableS2.docx