Computational and Structural Biotechnology Journal 23 (2024) 1418–1428

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Characteristics and phylogenetic distribution of megaplasmids and prediction of a putative chromid in *Pseudomonas aeruginosa*

Nanfei Wang [a,b,c,1], Xuan Zheng [d,1], Sebastian Leptihn [e,f,g], Yue Li [a,b,c], Heng Cai [a,b,c], Piaopiao Zhang [a,b,c], Wenhao Wu [a,b,c], Yunsong Yu [a,b,c,*], Xiaoting Hua [a,b,c,*]

[a] Department of Infectious Diseases, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China
[b] Key Laboratory of Microbial Technology and Bioinformatics of Zhejiang Province, Hangzhou, China
[c] Regional Medical Center for National Institute of Respiratory Diseases, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China
[d] Department of Nephrology, Sir Run Run Shaw Hospital, College of Medicine, Zhejiang University, Hangzhou, China
[e] HMU Health and Medical University, Am Anger 64/73 – 99084, Erfurt, Germany
[f] Deutsches Zentrum für Infektionsforschung (DZIF) Translational Phage-Network, Inhoffenstraße 7 – 38124, Braunschweig, Germany
[g] University of Southern Denmark,Department of Biochemistry and Molecular Biology, Campusvej 55 – 5230, Odense, Denmark

ARTICLE INFO

Keywords:
Pseudomonas aeruginosa
Antimicrobial resistance
Megaplasmid
Chromid
Extrachromosomal replicon

ABSTRACT

Research on megaplasmids that contribute to the spread of antimicrobial resistance (AMR) in *Pseudomonas aeruginosa* strains has grown in recent years due to the now widely used technologies allowing long-read sequencing. Here, we systematically analyzed distinct and consistent genetic characteristics of megaplasmids found in *P. aeruginosa*. Our data provide information on their phylogenetic distribution and hypotheses tracing the potential evolutionary paths of megaplasmids. Most of the megaplasmids we found belong to the IncP-2-type, with conserved and syntenic genetic backbones carrying modules of genes associated with chemotaxis apparatus, tellurite resistance and plasmid replication, segregation, and transmission. Extensively variable regions harbor abundant AMR genes, especially those encoding β-lactamases such as VIM-2, IMP-45, and KPC variants, which are high-risk elements in nosocomial infection. IncP-2 megaplasmids act as effective vehicles transmitting AMR genes to diverse regions. One evolutionary model of the origin of megaplasmids claims that chromids can develop from megaplasmids. These chromids have been characterized as an intermediate between a megaplasmid and a chromosome, also containing core genes that can be found on the chromosome but not on the megaplasmid. Using *in silico* prediction, we identified the "PABCH45 unnamed replicon" as a putative chromid in *P. aeruginosa*, which shows a much higher similarity and closer phylogenetic relationship to chromosomes than to megaplasmids while also encoding plasmid-like partition genes. We propose that such a chromid could facilitate genome expansion, allowing for more rapid adaptations to novel ecological niches or selective conditions, in comparison to megaplasmids.

1. Introduction

Pseudomonas aeruginosa is a major cause of nosocomial infections leading to high morbidity and mortality in cystic fibrosis patients or immunocompromised individuals. Eradication of *P. aeruginosa* is difficult due to its intrinsic and acquired drug resistance mechanisms [1]. Historically, studies showed a tendency to focus on resistance genes on various genomic islands and chromosomes but established a hypothesis that plasmids make a minor contribution to antimicrobial resistance (AMR) in *P. aeruginosa* strains [2]. With the extensive availability of long-read sequencing, however, a dramatic increase in the number of widespread *Pseudomonas* megaplasmid sequences that encode key traits of their host microorganisms has been observed recently [3]. The megaplasmids associated with *P. aeruginosa* (~300 to 500 kb) available in current literature are mostly classified as the incompatibility group P-2 (IncP-2) type, carrying cassette-borne carbapenemase genes in class 1 In/Tn with the conserved regions being involved in plasmid replication, maintenance, and conjugation [4–7].

According to George C et al. [8], bacterial replicons could be divided into chromosome and extrachromosomal replicon (second chromosome,

* Corresponding authors at: Department of Infectious Diseases, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China.
E-mail addresses: yvys119@zju.edu.cn (Y. Yu), xiaotinghua@zju.edu.cn (X. Hua).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2024.04.002
Received 8 October 2023; Received in revised form 1 April 2024; Accepted 1 April 2024
Available online 2 April 2024

chromid, megaplasmid, and plasmid). "Chromosome" refers to the primary replicon, the largest replicon with most of the core genes. "Second chromosome" is supposed to be split from an ancestral chromosome. "Megaplasmid and plasmid" are replicons carrying no core genes and the distinction between the two is mainly based on size. "Chromid", first described in 2010, is a novel concept of an in-between-element between a chromosome and a plasmid and acts as an independent type of replicon [9]. To some extent, chromid is a DNA chimera, a heterozygous molecule formed by the insertion of core genes from the chromosome into the megaplasmid. It is hypothesized that the long-term association of the megaplasmid with the coexistent chromosome eventually leads to a transfer of core genes under continued positive selection, resulting in a more stabilized "chromid" [3]. Such a genetic element allows more rapid evolutionary processes to adapt to novel niches, increasing fitness in the core biological networks of the genus [8]. Chromids possess several typical properties [10]. First, chromids have genomic characteristics that are more similar to those of chromosomes. Second, chromids carry several core genes that are essential for cell viability. Third, chromids have replication and maintenance machineries that more closely resemble those of plasmids.

In this study, we used all complete *P. aeruginosa* putative megaplasmid sequences that were available online and performed in-depth bioinformatic analyses comparing sequence content and structure. This information allowed us to examine their phylogenetic distribution to confirm the potential evolutionary relationship of these replicons. This study also attempts to predict the putative chromids of *P. aeruginosa in silico* for the first time.

2. Materials and methods

2.1. Analysis of bacterial genomes and identification of megaplasmids

All available 483 complete genome sequences of *P. aeruginosa* were obtained through the National Center for Biotechnology Information (NCBI) genome database on 15 November 2022. A lower cutoff of 300 kb was used to distinguish putative megaplasmid sequences of *P. aeruginosa*, according to the recommended literature [3]. Seventy-four complete genome sequences over 300 kb marked as "plasmid" were retrieved as potential megaplasmid sequences, and 45 corresponding chromosome sequences were available in the database (Table S1 and S2). The primary features of the hosts were retrieved manually from the Biosamples of the genomes uploaded to the NCBI. The countries labeled on the figure were the putative geographical sources. If the information of the country was lacking in the Biosample, we used the submitter's country in GenBank as the substitutive geographical source. The annotated "plasmid" sequences could be specifically divided into megaplasmids, putative chromids and secondary chromosomes based on the classification of a previous review [8]. Briefly, the extrachromosomal replicon with essential core genes was subdivided into a secondary chromosome if it resulted from a split of the chromosome into two. If not, it was defined as a putative chromid. The other secondary replicons carrying no core genes were defined as megaplasmids. Here, the term "extrachromosomal replicons" is used as a general term to refer to these 74 "plasmid" sequences in the following sections. The nucleotide sequences were reannotated via Prokka v1.14.6 [11]. ABRicate (https://github.com/tseemann/abricate) with the default database NCBI [12] and mlst v2.19.0 (https://github.com/tseemann/mlst) were used to identify resistance genes and multilocus sequence typing (MLST), respectively. The Average Nucleotide Identity (ANI) calculator was used to calculate the ANI value of prokaryotic genome sequences by the OrthoANIu algorithm [13]. The major use of ANI value in prokaryotic taxonomy is the demarcation of species, for which a cutoff of around 95–96% is often applied [14].

2.2. Phylogenetic distribution and genetic comparison of extrachromosomal replicons

The identification of core genes was established using the Panaroo v1.3.0 [15], and then a maximum-likelihood unrooted phylogenetic tree was constructed using IQ-TREE [16] and visualized using iTOL v6.1.1 [17]. The circular genome comparison of plasmids was performed using the BLAST Ring Image Generator BRIG v.0.95 [18]. Homology searches of plasmids were performed by BLASTN [19].

2.3. GC content and relative synonymous codon usage (RSCU) analysis

Potential chromids were identified on the basis of chromosome-like genetic features (GC content and RSCU), which were calculated by CodonW (http://codonw.sourceforge.net/). The RSCU values are close to 1.0 if all synonymous codons are used equally [20]. As the values for UGG (tryptophan) and AUG (methionine) are always 1.0, both were excluded as well as the three termination codons. Fifty-nine of the 64 possible codons were assessed through principal component analysis (PCA) calculated by GraphPad Prism v9.3.1. The distribution of the GC contents of every gene was visualized by ggplot2 package in R.

2.4. Identification and analysis of orthogroups

The inference of orthologous genes of the available 45 complete chromosome sequences and their corresponding 45 extrachromosomal replicons was performed by OrthoFinder [21,22], which inferred orthogroups, orthologues, the complete set of gene trees for all orthogroups and the rooted species tree. OrthoFinder was run with default settings [23]. In detail, GFF files of proteomes containing the amino acid sequences of genomes were uploaded and then all-vs-all BLAST comparisons of protein sequences with an e-value threshold of 10^{-3} were used to deduce putative phylogenetic relationships between pairs of genomes by reciprocal best similarity pairs. Then MAFFT [24] was used as the default multiple sequence alignment method and a maximum likelihood tree was inferred by FastTree [25]. Dendroscope v3.8.5 [26] was used to visualize the tree and the midpoint rooting method was used to root the tree. Each gene tree file obtained from OrthoFinder contained a phylogenetic tree based on the differences of a group of homologous genes that represent the evolutionary history of genes. We analyzed the proportion of homologous genes derived from megaplasmids in a gene tree in which a gene from an extrachromosomal replicon was located. All the gene trees containing the genes of that extrachromosomal replicon were taken into calculation and all the 45 extrachromosomal replicons were analyzed individually. The distribution of the total proportions was calculated by Perl scripts and visualized by ggplot2 package in R.

3. Results

3.1. Comparative genomic analyses and phylogenetic distribution of extrachromosomal replicons of P. aeruginosa

3.1.1. Comparative genomic analyses of extrachromosomal replicons of P. aeruginosa

Of the 74 complete extrachromosomal replicon sequences of *P. aeruginosa* (>300 kb) available through the NCBI genome database, we identified that the average and median genome sizes were ~444.2 kb and ~436.3 kb, with the PABCH45 unnamed replicon (923.2 kb) and the DN1 unnamed1 replicon (317.3 kb) being the maximum and minimum sizes, respectively. The earliest collection date of the strains was 1996, but most of the complete sequences were submitted after 2010 with the widespread use of long-read sequencing technologies. According to the analysis of Panaroo, a set of core genes (n = 113) was detected in 69 of the 74 replicons, with the exception of the replicons H15 unnamed, DN1 unnamed1, PA83 unnamed1, PABCH45 unnamed and 2021CK-01281

unnamed1. To further confirm whether the hosts of these 5 extrachromosomal replicons were *P. aeruginosa*, we used the Average Nucleotide Identity (ANI) calculator to calculate the ANI value of these prokaryotic genome sequences. We compared the chromosome sequences of these 5 extrachromosomal replicons to the chromosome of PAO1, the representative strain of *P. aeruginosa*. It demonstrated that the OrthoANIu values of these 5 comparisons were all over 98%, indicating that these 5 strains (H15, DN1, PA83, PABCH45 and 2021CK-01281) belonged to *P. aeruginosa*.

To further investigate the genetic structure of these complete replicon sequences, we first conducted a comparative analysis of the 69 megaplasmids sharing core genes (Fig. 1). The plasmid pOZ176 (accession no. KC543497) was used as a reference, as it had been classified as a member of the IncP-2 group by phenotypic incompatibility methods and the key features of the IncP-2 megaplasmids could be demonstrated when it was utilized as a reference. [27]. Since 51 of 69 replicons have been demonstrated to be IncP-2-type megaplasmids in previous studies (the references are shown in Table S1), an alignment of RepA protein encoded by *repP-2A* (pOZ176_183 gene, bp 109483 to 120876) of pOZ176 with the remaining 17 replicons indicated that all RepA proteins shared 99.51–100% amino acid sequence identity to that of pOZ176 (Fig. S1. A). The only exception was the plasmid pSE5369-VIM, which lacks the core genetic backbone containing the IncP-2-specific determinants (e.g., *repA*, *parAB*, *ter* and *che* operons) (Fig. S1. B). Thus, it can be concluded that 68 of 69 replicons belong to IncP-2-type megaplasmids, while the nature of the plasmid pSE5369-VIM has yet to be determined. The genetic comparison of the 68 IncP-2 megaplasmids demonstrated highly similar key traits, including genes encoding the replication and partition system (*repA* and *parAB*), a tellurite resistance operon (*terZABCDEF*) ubiquitously present in IncP-2-type plasmids, and *pil* operons involved in pilus assembly and twitching motility similar to the Pil-Chp system and chemotaxis operons (*CheBARZWY*). While the regions interpreted to be likely Integrative and Conjugative Elements (ICEs) in pOZ176 that contained the conjugative transfer modules (*trbK*, *traG*, and *trbBCDEJLFGI*) were absent in most of the megaplasmids [28], other conjugal transfer operons *traGBV*, *dnaG*, and type IV pilus/type II secretion system genes were found, which facilitated successful self-transmission in conjugation experiments[6,

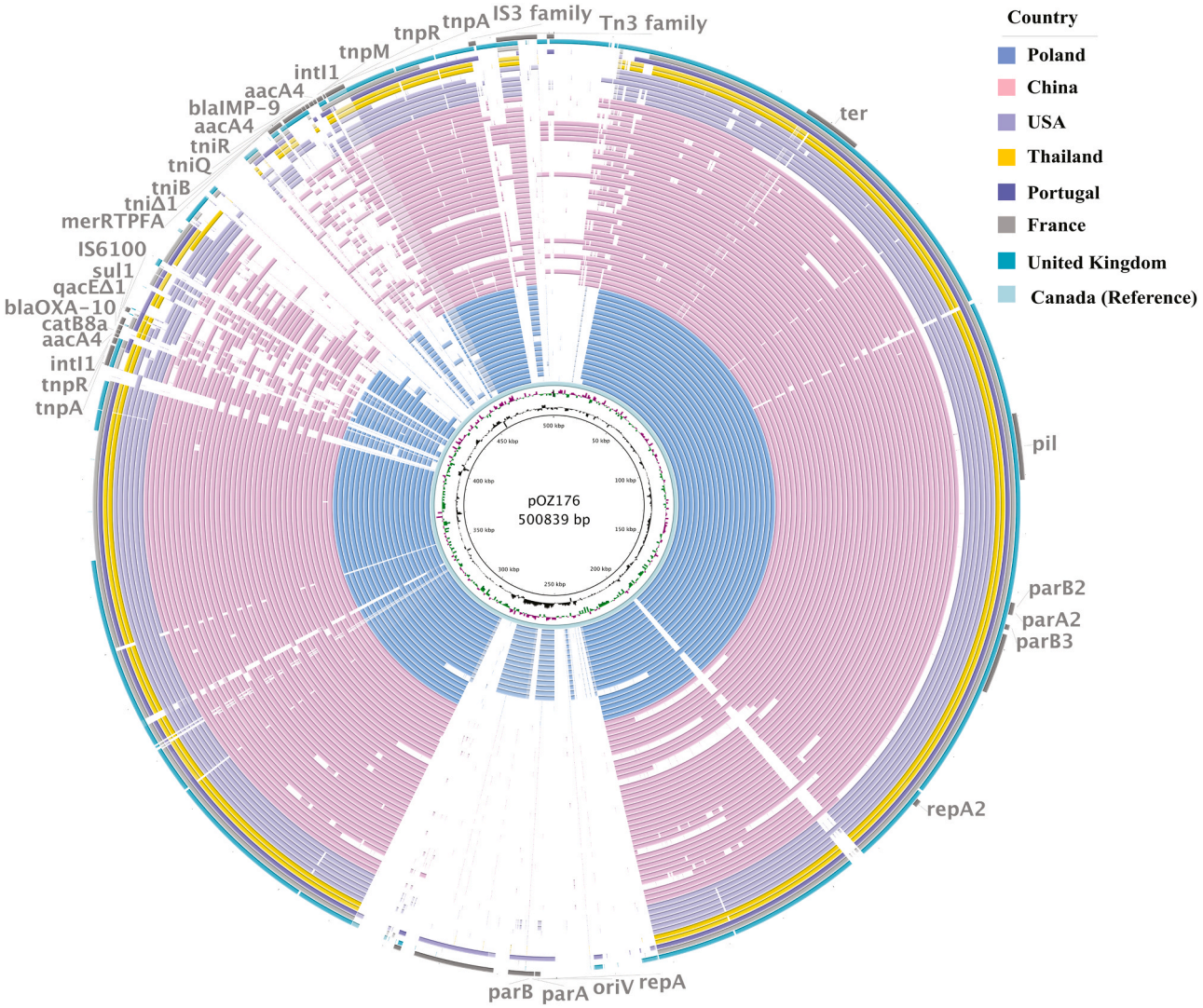

Fig. 1. Genetic comparison of 69 megaplasmids of *P. aeruginosa.* The comparison was generated using the BLAST Ring Image Generator (BRIG) v.0.95 and BLASTN. The GC content and GC skew were calculated by BRIG in the default parameter. The sequence of pOZ176 is taken as the reference and the locations of the main features of pOZ176 are annotated on the outermost circle. The solid regions demonstrate sequences similar to that of pOZ176, whereas the gaps represent regions lacking sequence similarity. The innermost circle indicates the scale, and the second and third circles represent the GC content and the GC skew, respectively. Each plasmid is colored and grouped based on the country it is isolated from and the information of the plasmids from the inner to the outer is illustrated in Table S1 (No.1-No.69).

29]. Apart from the conserved backbones, we observed that the accessory regions of megaplasmids varied based on resistance genes and related mobile genetic elements acquired during evolution.

We then performed homology searches of the remaining 5 replicons, which might represent "megaplasmids". The plasmids H15 unnamed and DN1 unnamed1 showed similarity to an IncP-9-type plasmid pKF715A (accession No. AP015030) sourced from *Pseudomonas putida* with 98.91% and 98.92% nucleotide identities, respectively. Compared with pKF715A, these two plasmids retained the regions encoding partitioning proteins and chemotaxis apparatus but lost the conjugative operon (*trwB*, *trwC* and type IV secretion system cluster), since the catabolic plasmid transferring from its environmental host to the clinical *P. aeruginosa* tended to delete the conjugative elements that might impose a large fitness cost on the new host. In addition, genetic modules related to the degradation of aromatic hydrocarbons, including the biphenyl and salicylate metabolism gene clusters (*bph-sal* element) and the benzoate catabolic gene (*bza*), were also lost (Fig. S2A) [30]. Most bacterial catabolic plasmids, such as IncP-9 plasmids carried by *P. putida*, are large (> 50 kb) and carry various genes that enable their host cells to utilize natural compounds and degrade aromatic hydrocarbons [31]. *P. putida* was mostly isolated from polluted environments due to its metabolic capacity and ability to acquire mobile genetic elements to adapt to specific environmental niches [30]. The megaplasmids might transfer from nonpathogenic *P. putida* to clinically common *P. aeruginosa* with a concurrent loss of the metabolic capacity to biodegrade toxic organics. Besides catabolic megaplasmids, the multidrug-resistant (MDR) megaplasmid pSY153-MDR carrying bla_{IMP-45} was also discovered in *P. putida* isolated from the urine of a cerebral infarction patient in China. This MDR megaplasmid was closely related to the other plasmids carried by *P. aeruginosa* isolated from China, supporting that between-species transmission of megaplasmids has occurred locally [32]. The plasmid PA83 unnamed1 was homologous to pMRCP2 of *Pseudomonas alcaligenes* (GenBank accession No. AP025274), exhibiting 93.81% nucleotide identity with intact replication, segregation and conjugation modules (Fig. S2B). The RepA protein of plasmid pA83 unnamed1 was highly similar to that of p35734-C [33] of the *Enterobacter cloacae* complex (accession No. CP010360), which belongs to the IncA/C group with 99.64% amino acid sequence identity.

3.1.2. *Phylogenetic distribution of 69 megaplasmids based on the nucleotide composition of core genes*

Based on the genetic composition of core genes, we constructed a phylogenetic tree of the 69 megaplasmids (Fig. 2). The megaplasmids were divided into two major groups. The majority of megaplasmids were found in strains isolated from China with a distribution in both *P. aeruginosa* clades. A cluster of the megaplasmids harboring bla_{VIM-2} genes was observed in one group, all of which were IncP-2-type strains isolated from strains prevalent in Polish nosocomial *P. aeruginosa* populations [6]. Most of the megaplasmids were found in *P. aeruginosa* strains isolated in different sources of patients in hospitals, such as the respiratory tract, digestive tract, skin wounds and infected areas. Only the megaplasmid pPA166–2-MDR with an array of AMR genes, which is closely related to the other megaplasmids isolated in China, was carried by *P. aeruginosa* strain PA166–2 isolated from the cloaca swab of a chicken in a poultry farm [34]. The strain belonged to ST313, which usually colonizes the intestines of healthy individuals but is rarely detected from the poultry environment. The available MLST of the associated strains was dispersed, with the most common type ST244 in five isolates, followed by ST463 and ST253.

To explore the role of plasmids in the transmission of drug-resistant

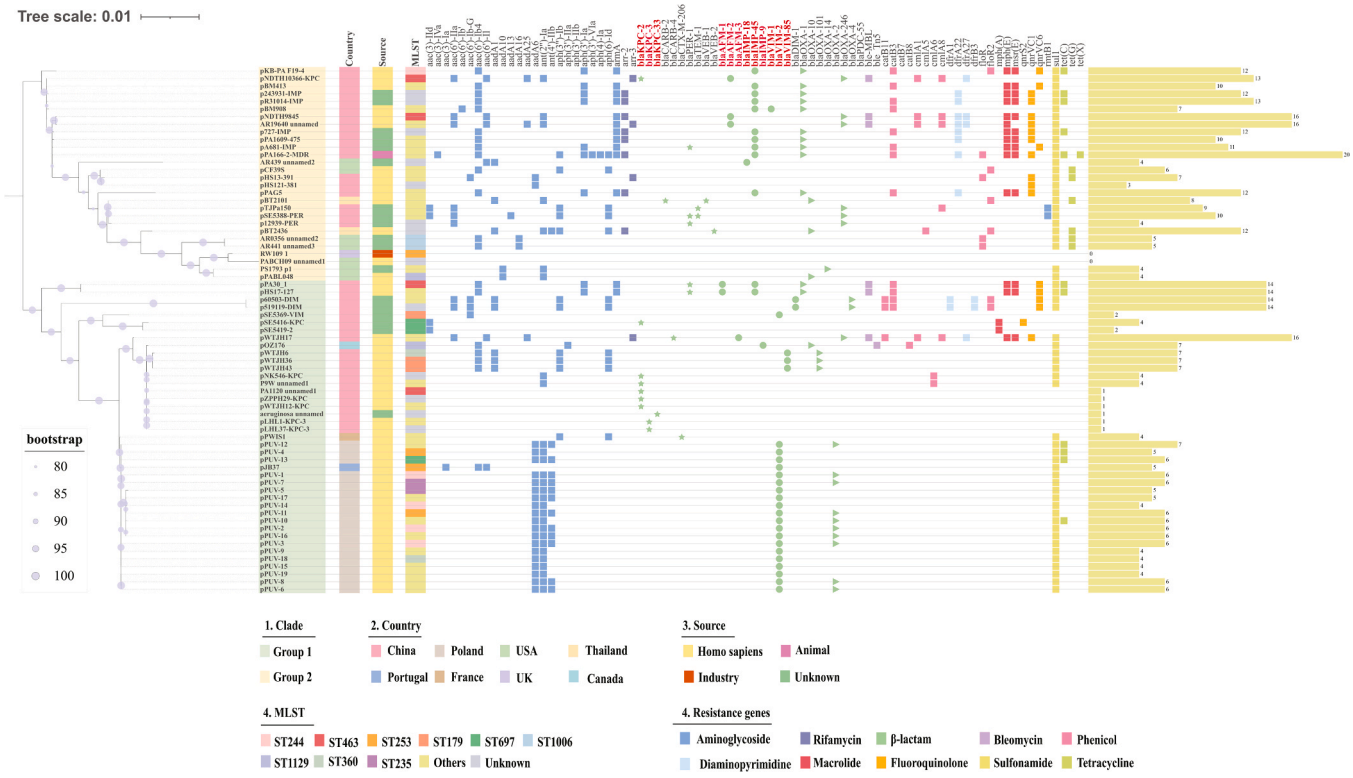

Fig. 2. Phylogenetic tree and the AMR gene content of 69 megaplasmids. The phylogenetic tree was constructed based on **a set of core genes** (n = 113). The core genome phylogeny was estimated from **the alignments of 113 core genes**. IQ-TREE was used for the calculation of the optimal base replacement model with the best-fit model GTR+F+I+G4 and the construction of the phylogenetic tree. The midpoint rooting method in iTOL v6.1.1 was used to distinguish the groups. The bootstrap values are indicated as the size of the blue dots on the spilt with the range of 80–100. The primary features of their hosts are indicated in various colored strips. The AMR genes are classified and colored based on the drug class they confer resistance to. The fields represent genes encoding different classes of β-lactamases are additionally distinguished in different shapes. The number of AMR genes in each megaplasmid is illustrated by a bar chart on the right.

genes, we grouped the AMR gene profiles in the clustering of the phylogenetic tree (Fig. 2). The AMR gene content of megaplasmids isolated from *P. aeruginosa* is extensive, with the dominant genes encoding aminoglycosidases and β-lactamases. We identified resistance genes of the classes A, B and D β-lactamases, in particular metallo-β-lactamases (MBLs), which had been reported as the epidemic territorial spread [5,6]. All the bla_{IMP-45} genes identified in the megaplasmids were located in the class 1 integron In*786* (*aacA4-*bla_{IMP-45}*-gcu35-*bla_{OXA-1}*-catB3*). Additionally, widespread evolutionary recruitment of resistance genes via IS*CR* modules was observed, e.g., IS*CR1-armA*, IS*CR1-qnrVC6*, IS*CR1-*bla_{PER-1} and IS*CR27n3-*bla_{AFM-1}. Variants of the bla_{AFM} genes bla_{AFM-2} and bla_{AFM-3} were also embedded in IS*CR* units adjacent to Tn*1403*-derived integrons [35] (Fig. S3). Another common type of MBLs gene was bla_{VIM-2}, primarily found in In*461* with a cassette array of *aadB-*bla_{VIM-2}*-aadA6* in the pPUVs megaplasmids [6]. All the bla_{KPC} genes, including bla_{KPC-2}, bla_{KPC-3} and bla_{KPC-33}, were isolated in China. The majority of the variants of bla_{KPC} were transmitted as part of a conserved genetic platform IS*6100*-IS*Kpn27-*bla_{KPC}-ΔIS*Kpn6-korC-klcA* in Tn*6296*, while pNDTH10366-KPC contained two copies of the bla_{KPC-2} genes as the result of the inversion and duplication of the IS*26-*bla_{KPC-2}-IS*26* unit [36] (Fig. S4). Nevertheless, two of the megaplasmids, plasmid unnamed1 of PABCH09 and plasmid 1 of RW109 (the largest industrial *P. aeruginosa* strain in the dataset) [37], lacked any AMR genes and grouped together in a subgroup (Fig. 2).

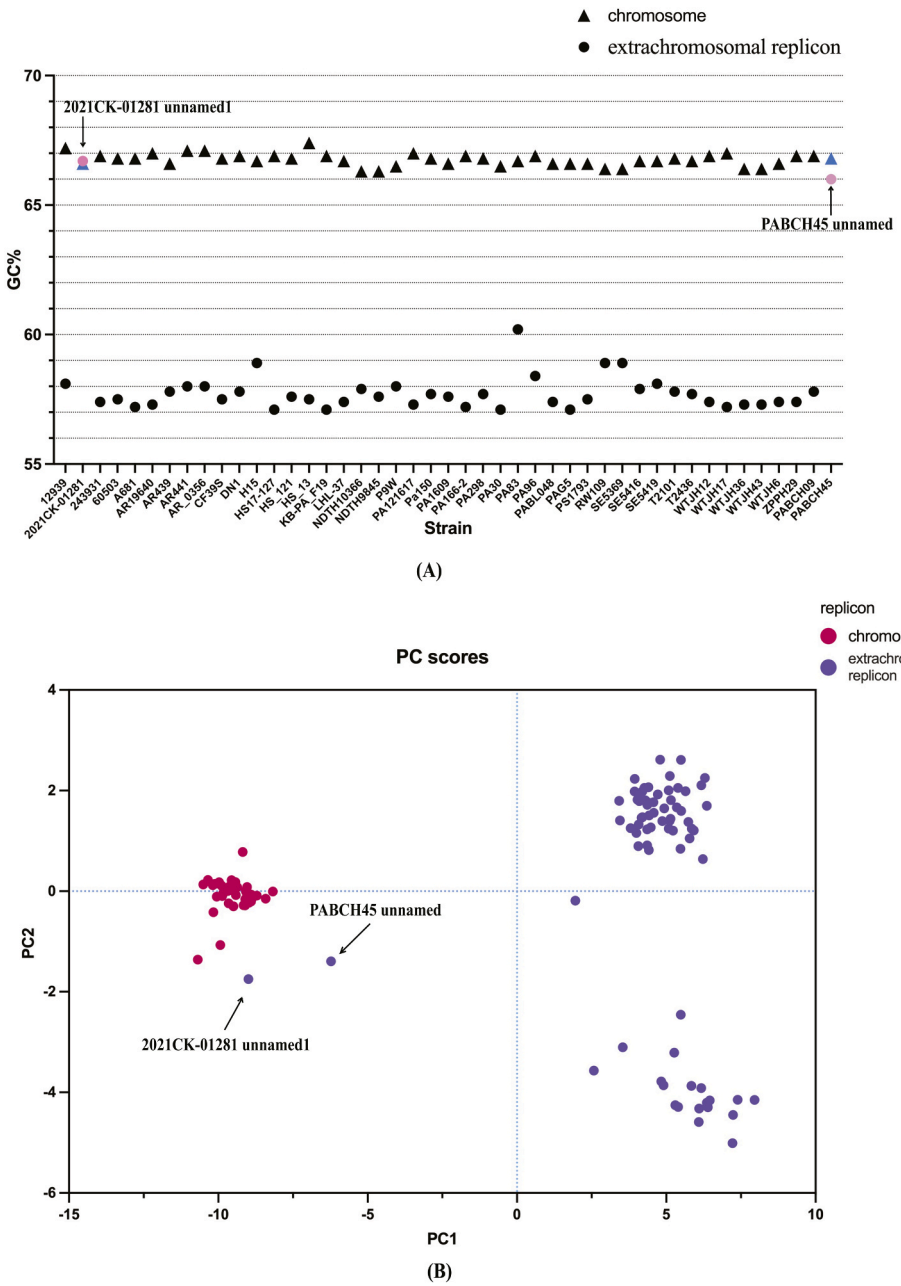

Fig. 3. The comparison of GC content and RSCU of each extrachromosomal replicon with the corresponding 45 chromosomes. **(A).** The difference in GC content and the distributions are presented for chromosomes (solid triangle) and extrachromosomal replicons (solid circle). The putative chromids and corresponding chromosomes are highlighted in blue and pink, respectively. **(B).** Principal components plot of RSCU for chromosomes (red dot) and extrachromosomal replicons (purple dot). The proportion of variance of Principal components 1 and 2 are 84.47% and 5.92%, respectively. The GC content and RSCU value of each replicon are listed in Table S2.

3.2. Prediction of putative chromids

3.2.1. Genomic signatures of the replicons PABCH45 unnamed and 2021CK-01281 unnamed1

Homology searches in the NCBI database did not identify any plasmid that was similar to the ~0.92-Mb replicon of "PABCH45 unnamed" or the ~0.59-Mb replicon of "2021CK-01281 unnamed1″. We analyzed the genomic features of the two replicons, including their GC content (percentage of the genome consisting of guanine-cytosine) and their relative synonymous codon usage (RSCU, the ratio of the observed frequency of a specific codon to the expected value). The analysis of these genomic features of extrachromosomal replicons helps distinguish the replicon types (megaplasmid and/or putative chromid). The analyses were based on the available 45 complete chromosome sequences and their corresponding extrachromosomal replicons, including untyped "PABCH45 unnamed" and "2021CK-01281 unnamed1″ replicons as well as 43 of the megaplasmids mentioned above (Table S2).

Our analysis revealed that all 43 megaplasmids had a lower GC content than that of the chromosomes averaging 58% and 67%, respectively. In contrast, the "PABCH45 unnamed" and the "2021CK-01281 unnamed1″ replicons had a much similar GC content when compared to the chromosomes found in the same strains, differing by 0.8% and 0.1%, respectively (Fig. 3A). Not only did the extent of differences in GC content differ between "PABCH45 unnamed" and "2021CK-01281 unnamed1″ and the megaplasmids, but the deviation of the RSCU values from that of the host chromosomes also appeared to be clearly different. The RSCU values for "PABCH45 unnamed" and "2021CK-01281 unnamed1″ were very similar to those of the chromosomes but clearly distinct from those of the 43 megaplasmids (Fig. 3B). It indicated that these genetic differences between the chromosomes and the two replicons ("PABCH45 unnamed" and "2021CK-01281 unnamed1″) were much less than the differences between the chromosomes and the megaplasmids.

Additionally, we calculated the GC contents of every individual gene of 45 replicons, each of which contains a chromosome and its corresponding extrachromosomal replicon. It showed that the difference in the total GC contents of all replicons was not significant, mainly ranging from 0.65–0.7 (Fig. 4A), and neither did the GC contents of every core gene (Fig. 4B). It might be that the conserved core genes derived from chromosomes were in the majority of the core genes of all the included replicons, which made it difficult to distinguish the GC contents of extrachromosomal replicons, especially the GC contents of putative chromids. We further calculated the individual non-core genes of all the 74 extrachromosomal replicons, and it demonstrated that the GC contents of" PABCH45 unnamed" and "2021CK-01281 unnamed1″ were significantly higher than those of the other megaplasmids (0.65–0.7 vs 0.55–0.6), indicating that the genomic signatures of the replicons "PABCH45 unnamed" and "2021CK-01281 unnamed1″ were deviated from the megaplasmids (Fig. 4C). However, when we calculated the RSCU for every core gene and non-core gene of the replicons mentioned above, neither the RSCU values for every core gene of extrachromosomal replicons were distinguished from those of chromosomes, nor the RSCU values for non-core genes.

3.2.2. Identification and analysis of orthogroups

We performed an analysis of the homologues contained within the 45 genomes from the complete chromosome of *P. aeruginosa* strains and the extrachromosomal replicon sequences. The overlaps of homologues identified within genomes are shown in Fig. S5. Within the heatmap, three regions can be observed (Regions A, B and C) according to the pairwise relationships within replicons. A comparison of Region A and Region B illustrated that the homologous genes in the chromosome were much more numerous than those in the extrachromosomal replicons since genes in chromosomes were transmitted relatively stably by

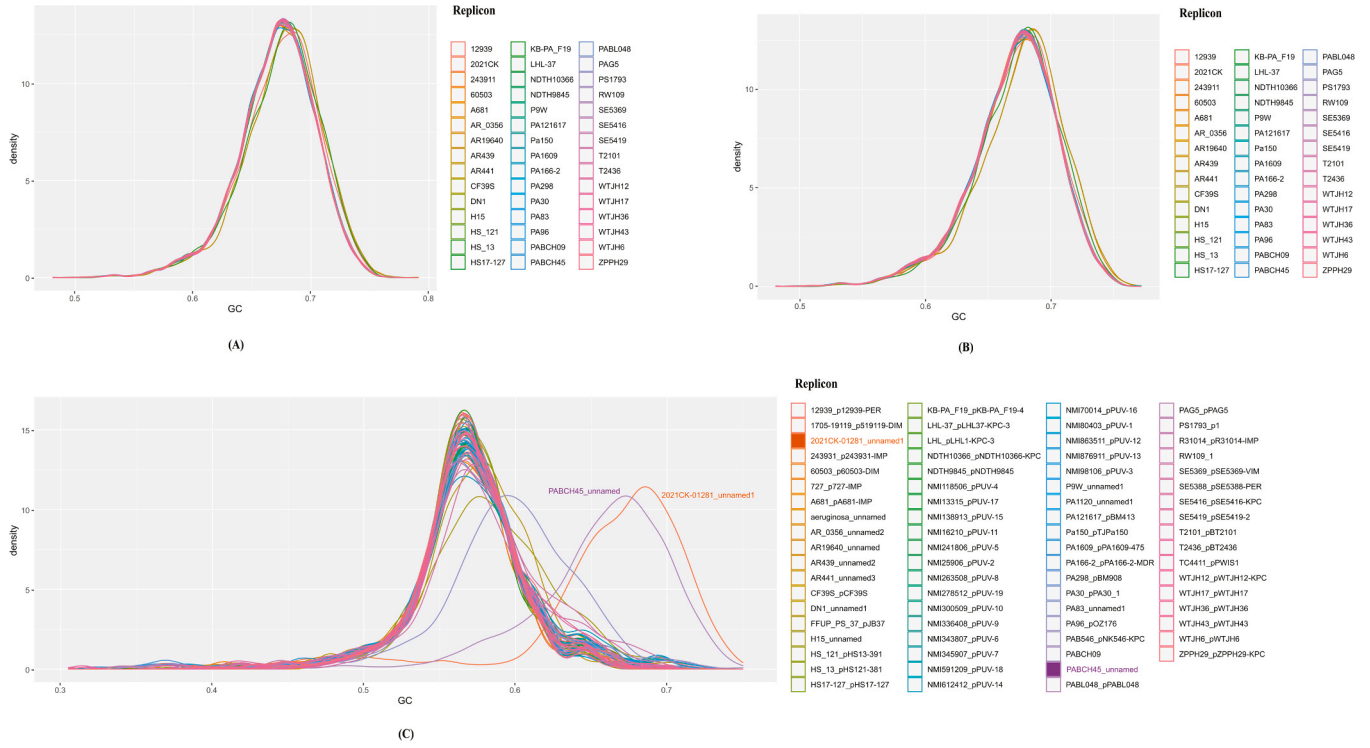

Fig. 4. The GC content of individual genes, including both core and non-core accessory genes. **(A)** The GC contents of every individual gene of 45 replicons. **(B)** The GC contents of every core gene of 45 replicons. **(C)** The GC contents of every non-core gene of 74 extrachromosomal replicons. The GFF files of the 45 chromosome sequences and their corresponding 45 extrachromosomal replicons were analyzed by Panaroo. The FASTA files of the aligned gene sequences produced by Panaroo were uploaded to CodonW to calculate the GC contents of every core gene and the rest non-core genes were also calculated. The distribution of the GC contents of every gene was visualized by ggplot2.

vertical inheritance. The pairwise comparisons between chromosomes and extrachromosomal replicons showed that the grids corresponding to the "PABCH45 unnamed" and "2021CK-01281 unnamed1″ replicons were displayed darker than others, illustrating partially higher similarity of genes between these two replicons and the chromosomes. Moreover, the phylogenetic tree based on orthogroups assigns the "PABCH45 unnamed" and "2021CK-01281 unnamed1″ replicons within the chromosomal replicons separated from the group represented by the megaplasmids (Fig. 5).

To evaluate the phylogenetic relationships within individual genes of the 45 chromosome genomes and corresponding 45 extrachromosomal replicons, we analyzed the proportion of homologous genes derived from megaplasmids in a gene tree in which a gene from an extrachromosomal replicon was located. The higher the ratio, the more related the extrachromosomal replicon gene was to the megaplasmids. Conversely, the smaller ratio indicated that the proportion of homologous genes derived from chromosomes increased accordingly, suggesting that the genes of this extrachromosomal replicon were more related to chromosome genes and had a closer evolutionary relationship with chromosomes. The result demonstrated that the genes in "2021CK-02381_unnamed1″ and "PABCH45_unanmed" (shown in the orange rectangles in Fig. 6) were more homologous with the genes of chromosomes. In contrast, other replicons verified as megaplasmids before showed opposite results as predicted. It was further proven that "2021CK-02381 unnamed1″ and "PABCH45 unnamed" were potential putative chromids.

3.2.3. *Phylogenetic analysis of the partitioning protein ParA*

We further analyzed a phylogenetic tree of the protein that is essential for plasmid partition, ParA. The sequences encoded by the megaplasmids and the replicons "PABCH45 unnamed" and "2021CK-01281 unnamed1″ would provide clues as to whether the two replicons

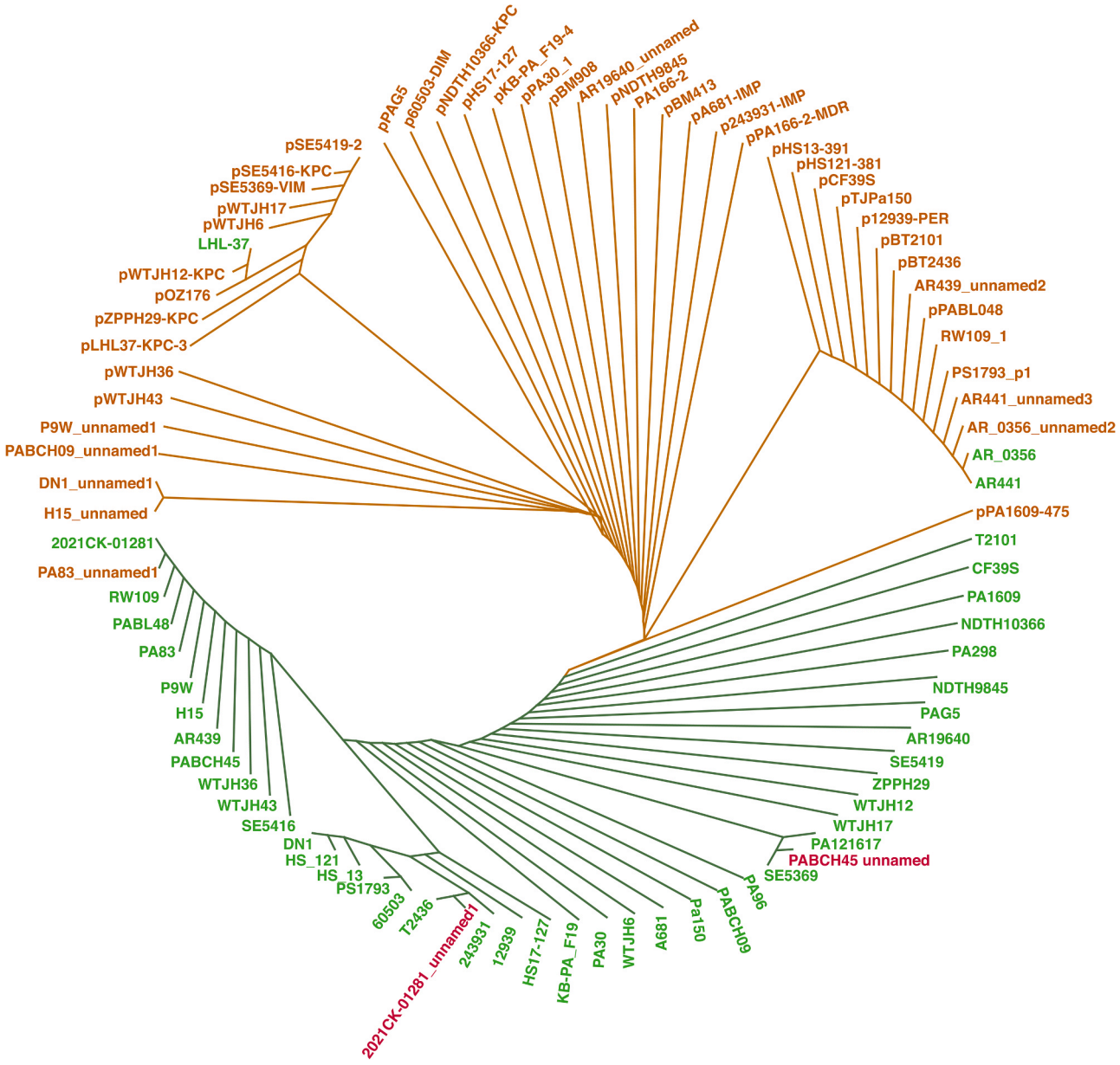

Fig. 5. Phylogenetic tree based on orthogroups of 45 extrachromosomal replicons with the corresponding chromosomes. The maximum-likelihood unrooted phylogenetic tree of orthogroup is constructed by OrthoFinder and FastTree. The colors of the branches represent the groups of the replicons. The labels in brown represent the extrachromosomal replicons and the labels in green represent the chromosomal replicons. The labels of extrachromosomal replicons "PABCH45 unnamed" and "2021CK-01281 unnamed1″ were highlighted in red.

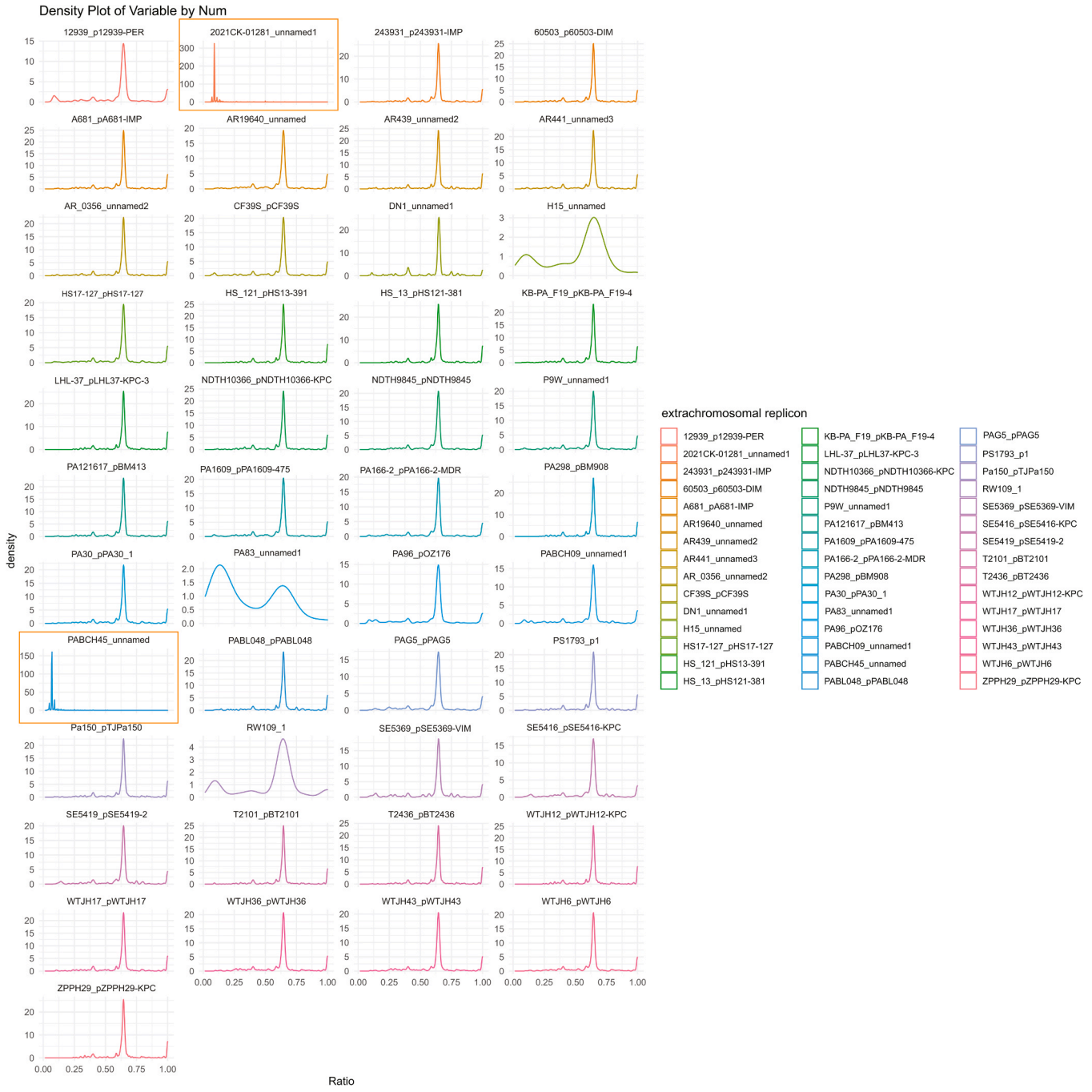

Fig. 6. Analysis of each phylogenetic tree of groups of homologous genes of 45 chromosome genomes and corresponding 45 extrachromosomal replicons. The distribution of the total proportions was calculated by Perl scripts and visualized by ggplot2. The distributions of the proportions in "2021CK-02381 unnamed1″ and "PABCH45 unnamed" were highlighted in orange rectangles.

might be designated as potential chromids of *P. aeruginosa*. The phylogeny of partitioning systems showed that all megaplasmids had closely related partitioning proteins, with the "PABCH45 unnamed" encoded ParA protein homologous to that of the megaplasmids. However, the "2021CK-01281 unnamed1″ partitioning protein grouped separately (Fig. S6). Additionally, synteny analysis of "2021CK-01281 unnamed1″ revealed high consistency with the chromosomes of PAO1, the representative strain of *P. aeruginosa* (Fig. S7).

4. Discussion

Megaplasmids, as the name suggests, are very large plasmids that were first described by Rosenberg in the early 1980 s [38]. Thresholds for minimum megaplasmid sizes have not yet been established, as genome size itself varies greatly across phyla. DiCenzo et al. suggested a lower cutoff of 350 kb for megaplasmids, as this was 10% of the median bacterial genome size [8], while J. Hall et al. noted that fixed thresholds tend to bias megaplasmids toward species with large genomes. The latter group considered it more appropriate to set the threshold of megaplasmid based on 5% of the total genome size of that species, referring to megaplasmid as a relatively large plasmid to the genome it comes from [3]. In our work, we used a lower cutoff of 300 kb to define putative megaplasmids of *P. aeruginosa*, correlating to 5% of the genome size of this species and corresponding to the majority of the megaplasmid size

reported in *P. aeruginosa* isolates. However, distinguishing megaplasmids based solely on size is arbitrary since the distribution of plasmid sizes can vary widely. A more functional, possibly sequence-based definition of "megaplasmids" could be considered in the future.

IncP-2-type megaplasmids, a large group of *Pseudomonas* megaplasmids carrying AMR genes, were defined by experimental incompatibility in the 1970 s [39]. Currently, the prediction of a novel class of plasmids based on the alignment of amino acid sequences of the RepA proteins is possible due to the development of long-read sequencing technology and comparison possibilities due to the massive expansion of the plasmid databases. According to our comprehensive analysis of 74 complete extrachromosomal replicon sequences of *P. aeruginosa*, 68 replicons were verified as IncP-2 group megaplasmids, including 17 replicons first confirmed in our report. Our genomic analysis confirmed the conserved and syntenic core backbone of the IncP-2 megaplasmid family, which carries genes encoding chemotaxis apparatus, a type IV pilus, and plasmid replication, segregation, and transmission. However, the majority of members possess extensively variable regions with abundant AMR genes, especially those encoding β-lactamases in nosocomial isolates. Urbanowicz et al. [6] reported the IncP-2 megaplasmids as AMR/MDR platforms carrying cassette-borne $bla_{\text{VIM-2}}$ genes in In*461* disseminating in nosocomial *P. aeruginosa* populations in Poland. Zhang et al. [5] reported outbreaks of IMP-45-producing *P. aeruginosa* in China, which was attributed to the worldwide spread of IncP-2 megaplasmids. It should be noted that IncP-2 megaplasmids harboring $bla_{\text{KPC-2}}$ genes, the carbapenem resistance genes mainly carried by type I plasmids with the core genetic platform IS*Kpn27*-$bla_{\text{KPC-2}}$-IS*Kpn6* [36], are frequently identified across China. KPC-producing *P. aeruginosa* strains are the predominant carbapenem-resistant *P. aeruginosa* (CRPA) strains in China, and IncP-2 megaplasmids may serve as mediators facilitating the dissemination of such resistance genes. These current observations may illustrate the geographical and nosocomial distribution of the IncP-2 megaplasmids from *P. aeruginosa*. However, since the available complete genomic data is still insufficient, the majority of the available long-read sequenced genomes were isolated from China and Poland mainly due to the supervision of the epidemic territorial dissemination of the carbapenemase-encoding megaplasmids in the *P. aeruginosa* population. With the development of the long-read sequencing technology, a more detailed profile of the distribution of IncP-2 megaplasmids might be built up.

While nosocomial isolates are the most common sources, one megaplasmid (pPA166–2-MDR) with multiple AMR genes from chickens was obtained, which is closely related to the other megaplasmids isolated in clinical settings. This suggests that the poultry could have been contaminated by human activities and the IncP-2 megaplasmid of *P. aeruginosa* harboring various resistance genes has spread between humans and animals. Additionally, the high stability, low fitness cost, and efficient transferability as well as the conserved ancestral backbone of IncP-2 megaplasmids make them effective vehicles of AMR gene circulation in diverse hosts [4,5,29]. However, the IncP-2 plasmids are generally underrepresented in veterinary isolates because of the low frequency of the detection of MDR plasmids in the poultry or veterinary industry as well as the less available long-read sequencing data.

Aside from megaplasmids, the term "chromid" was introduced to describe replicons with properties of both chromosomes and plasmids [9]. Evolution results in the optimization of the genetic material that also shapes the nucleotide composition of the chromids, resulting in the similarity of the genomic structures between bacterial chromosomes and associated chromids. From a practical perspective, chromids are typically uploaded to plasmid databases as a type of large plasmid and are not defined as such [40]. This is also because it is less clear when a sequence is a plasmid, a megaplasmid or a chromid. The *in silico* differentiation of putative chromids and megaplasmids is mainly based on two characteristics: first, the chromids also contain genes coding for proteins that are essential for cell viability; second, their nucleotide composition, such as the GC content and synonymous codon usage, more closely resembles that of the chromosomes [9].

GC content can not only indicate genes recently acquired through horizontal transfer but also vary widely within each replicon in a genome, with the degree of variation being reflective of the replicon forms [8,41]. Previously, cutoff values of not more than 1% or 2% GC difference of a putative chromid compared to the bacterial chromosome were proposed [8,10]. The comparatively lower GC content of nonessential self-replicating genetic elements, such as plasmids, is the result of the selection of functional and ecological requirements, concurrent with energy optimization, as GTP and CTP nucleotides have higher energy expenditure [42]. Putative chromids with higher GC content might thus be identified, as the identical selection pressure is acting on them and chromosomes. The RSCU value is a frequently used indicator to measure codon usage bias (CUB) [43,44]. CUB is ubiquitous in genomes and varies frequently across species [45]. Henry I *et al.* stated that codon usage bias could help in estimating gene expression levels [46], which was also supported by Tessa E.F.Quax *et al.*[47]. The RSCU of the putative chromid is much closer to that of the chromosome compared to the plasmid, indicating a longer evolutionary coexistence in the same cellular environment and a closer genetic relationship between the chromosome and the co-residing chromid. In our study, the GC content and RSCU values of the "PABCH45 unnamed" and "2021CK-01281 unnamed1″ replicons were distinctly more similar to the chromosomes than plasmids, indicating that we putatively identified two chromids in *P. aeruginosa* strains. However, according to Haruo Suzuki et al.[48], although PCA has often been used to identify major trends of variation in RSCU among genes, it also has known biases which were reported to be affected by a bias associated with the rarity of cysteine in the protein.

In addition to the GC content and RSCU values, the existence of orthologous gene pairs within the chromosome and the correlating extrachromosomal replicon provides further evidence that the elements are indeed chromids [10]. Orthologs are genes that originate from a common ancestor through a speciation event, resulting in similar genes in different species, which also allows for functional and evolutionary assessments of the replicons to each other. In our study, the phylogenetic analysis suggested that the "PABCH45 unnamed" and "2021CK-01281 unnamed1″ replicons are much closer to the chromosomes than the megaplasmids. In addition, a large number of orthogroups were found in the putative chromids that have their corresponding counterparts in the chromosomes; for example, "PABCH45 unnamed" shares 4815 orthogroups with chromosomes PA121617 and SE5369 (Fig. S8). More precisely, the genes flow from the chromosome to the extrachromosomal replicons resulting in homologous genes in chromids encoding core functional features. The conserved set of core essential genes service the fundamental processes such as protein synthesis and information transfer, including not only known housekeeping genes essential for growth under all conditions but also genes whose function and essentiality are poorly understood [9]. However, the definition of the core genes is currently difficult to determine and there are no genes that are universally chromid-encoded. Qing-Hua Zou et al. [49] defined core genes as the common orthologs in all compared genomes and Nicola Segata et al. [50] referred core genes to one or more gene families highly conserved at the nucleotide sequence level of a related genome group.

Previous phylogenetic analysis of the chromid partitioning proteins ParAB in *Alphaproteobacteria* demonstrated that all chromids within a genus possess closely related plasmid-like replication and partition genes [9]. In our study, the putative chromid "PABCH45 unnamed" has a plasmid-like ParA protein, whereas the ParA protein of the other replicon "2021CK-01281 unnamed1″ is distinctly separate, outside of the group of plasmid partitioning systems. It may even possibly be appropriate to characterize such a "2021CK-01281 unnamed1″ replicon as a second chromosome rather than as a chromid, since further synteny analysis of the replicon indicates significant genomic synteny with the chromosome of PAO1, inferring the formation of this replicon

originating from a separation event of an ancestral chromosome. However, second chromosomes are rare in bacterial genomic datasets, and it may be problematic to distinguish a second chromosome from chromids in large data analyses[8].

Putative chromids have thus far been found in several genera associated with eukaryotic organisms in symbiotic or pathogenic relationships [8]. One of the most fascinating and widely accepted evolutionary hypotheses is that the transfer of megaplasmids into the host by horizontal gene transfer (HGT) allows the exploitation of new ecological niches, which in turn drives the continuous coevolution of the chromosome and the coresident megaplasmid. This often results in the transfer of core genes from the chromosome to the megaplasmid as well as the loss of horizontal transmission of the megaplasmid, which facilitates the formation of a chromid [3,8]. Under these circumstances, the coordinated regulation of genetic elements would allow for better adaptation to novel ecological niches compared to unstable and more dynamic megaplasmids. Other evolutionary advantages of chromids are that they allow genome expansion due to stability (as they are retained in the host) and faster bacterial division, as each replicon of the genome can replicate independently and in parallel [51,52]. Putative chromids were identified mainly based on bioinformatic tools in previous studies; however, to verify the validity of the conclusion that one replicon is indeed a chromid, it is essential to perform in vitro experiments e.g., the target-oriented replicon curing technique or mutational analysis [10], which would also allow us to gain more knowledge of this genomic module.

5. Conclusion

In our work, we systematically analyzed the distinct and consistent genetic characteristics of megaplasmids found in *P. aeruginosa*. We provide their phylogenetic distribution and predict - for the first time - a putative chromid using a combination of *in silico* approaches while also presenting potential evolutionary paths of megaplasmids in *P. aeruginosa*.

Research data for this article

The data that support the findings of this study are available in Mendeley at https://data.mendeley.com/datasets/8g3krbjpst/2 (doi: 10.17632/8g3krbjpst.2).

CRediT authorship contribution statement

Nanfei Wang: Data curation, Writing – original draft. **Xuan Zheng:** Writing – original draft. **Sebastian Leptihn:** Writing – review & editing. **Yue Li:** Data curation. **Heng Cai:** Supervision, Writing – review & editing. **Piaopiao Zhang:** Supervision, Writing – review & editing. **Wenhao Wu:** Data curation. **Yunsong Yu:** Conceptualization. **Xiaoting Hua:** Conceptualization, Writing – review & editing.

Declaration of Competing Interest

None.

Acknowledgments

This work was supported by grants from the National Key Research and Development Program (No. 2018YFE0101800) and the Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang Province (No. 2021R01012) awarded to YY, and the Natural Science Foundation of Zhejiang Province (No. LZY24H150003) awarded to XH. The graphical abstract was created with BioRender.com. The PacBio raw reads and HGAP assembly files of PABCH45 were kindly provided by Dr. Gregory P Priebe.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.04.002.

References

[1] Pang Z, Raudonis R, Glick BR, Lin TJ, Cheng Z. Antibiotic resistance in Pseudomonas aeruginosa: mechanisms and alternative therapeutic strategies. Biotechnol Adv 2019;37(1):177–92 (Feb).
[2] Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. Front Microbiol 2015;6:242.
[3] Hall JPJ, Botelho J, Cazares A, Baltrus DA. What makes a megaplasmid? Philos Trans R Soc B 2022;377(1842):20200472. Jan 17.
[4] Cazares A, Moore MP, Hall JPJ, Wright LL, Grimes M, Emond-Rhéault JG, et al. A megaplasmid family driving dissemination of multidrug resistance in Pseudomonas. Nat Commun 2020;11(1):1370 (Dec).
[5] Zhang X, Wang L, Li D, Li P, Yuan L, Yang F, et al. An IncP-2 plasmid sublineage associated with dissemination of *bla* $_{IMP-45}$ among carbapenem-resistant *Pseudomonas aeruginosa*. Emerg Microbes Infect 2021;10(1):442–9. Jan 1.
[6] Urbanowicz P, Bitar I, Izdebski R, Baraniak A, Literacka E, Hrabák J, et al. Epidemic territorial spread of IncP-2-Type VIM-2 carbapenemase-encoding megaplasmids in nosocomial Pseudomonas aeruginosa populations. Antimicrob Agents Chemother 2021;65(4):e02122-20. Mar 18.
[7] Fang Y, Wang N, Wu Z, Zhu Y, Ma Y, Li Y, et al. An XDR Pseudomonas aeruginosa ST463 strain with an IncP-2 plasmid containing a novel transposon Tn6485f encoding blaIMP-45 and blaAFM-1 and a second plasmid with two copies of blaKPC-2. Microbiol Spectr 2023;11(1):e0446222. Feb 14.
[8] diCenzo GC, Finan TM. The divided bacterial genome: structure, function, and evolution. Microbiol Mol Biol Rev 2017;81(3):e00019-17 (Sep).
[9] Harrison PW, Lower RPJ, Kim NKD, Young JPW. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. Trends Microbiol 2010;18(4):141–8 (Apr).
[10] Dziewit L, Bartosik D. Comparative analyses of extrachromosomal bacterial replicons, identification of chromids, and experimental evaluation of their indispensability. Methods Mol Biol 2015;1231:15–29.
[11] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30 (14):2068–9. Jul 15.
[12] Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. Antimicrob Agents Chemother 2019;63(11):e00483-19 (Nov).
[13] Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek 2017;110(10): 1281–6 (Oct).
[14] Chun J, Rainey FA. Integrating genomics into the taxonomy and systematics of the bacteria and archaea. Int J Syst Evolut Microbiol 2014;64(Pt_2):316–24.
[15] Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol 2020;21(1):180. Jul 22.
[16] Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 2015;32(1):268–74 (Jan).
[17] Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 2019;47(W1):W256–9. Jul 2.
[18] Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genom 2011;12:402. Aug 8.
[19] Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, et al. BLAST: a more efficient report with usability improvements. Nucleic Acids Res 2013;41(Web Server issue):W29–33 (Jul).
[20] Gupta SK, Bhattacharyya TK, Ghosh TC. Synonymous codon usage in Lactococcus lactis: mutational bias versus translational selection. J Biomol Struct Dyn 2004;21 (4):527–36 (Feb).
[21] Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 2019;20(1):238. Nov 14.
[22] Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 2015;16(1):157. Aug 6.
[23] Kelly S, Maini PK. DendroBLAST: approximate phylogenetic trees in the absence of multiple sequence alignments. PLoS One 2013;8(3):e58537.
[24] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30(4):772–80 (Apr).
[25] Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 2010;5(3):e9490. Mar 10.
[26] Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol 2012;61(6):1061–7. Dec 1.
[27] Sagai H, Hasuda K, Iyobe S, Bryan LE, Holloway BW, Mitsuhashi S. Classification of R plasmids by incompatibility in Pseudomonas aeruginosa. Antimicrob Agents Chemother 1976;10(4):573–8 (Oct).
[28] Jiang X, Yin Z, Yuan M, Cheng Q, Hu L, Xu Y, et al. Plasmids of novel incompatibility group IncpRBL16 from Pseudomonas species. J Antimicrob Chemother 2020;75(8):2093–100. Aug 1.

[29] Li M, Guan C, Song G, Gao X, Yang W, Wang T, et al. Characterization of a conjugative multidrug resistance IncP-2 megaplasmid, pPAG5, from a clinical Pseudomonas aeruginosa isolate. Khursigara CM, editor. Microbiol Spectr 2022. Feb 23;10(1):e01992-21.
[30] Suenaga H, Fujihara H, Kimura N, Hirose J, Watanabe T, Futagami T, et al. Insights into the genomic plasticity of Pseudomonas putida KF715, a strain with unique biphenyl-utilizing activity and genome instability properties. Environ Microbiol Rep 2017;9(5):589–98.
[31] Sota M, Yano H, Ono A, Miyazaki R, Ishii H, Genka H, et al. Genomic and functional analysis of the IncP-9 naphthalene-catabolic plasmid NAH7 and its transposon Tn4655 suggests catabolic gene spread by a tyrosine recombinase. J Bacteriol 2006;188(11):4057–67.
[32] Yuan M, Chen H, Zhu X, Feng J, Zhan Z, Zhang D, et al. pSY153-MDR, a p12969-DIM-related mega plasmid carrying *bla* IMP-45 and *armA*, from clinical *Pseudomonas putida*. Oncotarget 2017;8(40):68439–47. Sep 15.
[33] Chavda KD, Chen L, Fouts DE, Sutton G, Brinkac L, Jenkins SG, et al. Comprehensive genome analysis of carbapenemase-producing Enterobacter spp.: new insights into phylogeny, population structure, and resistance mechanisms. mBio 2016;7(6):e02093-16. Dec 13.
[34] Dong N, Liu C, Hu Y, Lu J, Zeng Y, Chen G, et al. Emergence of an extensive drug resistant Pseudomonas aeruginosa strain of chicken origin carrying blaIMP-45, tet (X6), and tmexCD3-toprJ3 on an IncpRBL16 plasmid. Microbiol Spectr 2022;10(6): e0228322. Dec 21.
[35] Li Y, Zhu Y, Zhou W, Chen Z, Moran RA, Ke H, et al. Alcaligenes faecalis metallo-β-lactamase in extensively drug-resistant Pseudomonas aeruginosa isolates. Clin Microbiol Infect 2022;28(6):880.e1–8.
[36] Zhu Y., Chen J., Shen H., Chen Z., Yang Q. wen, Zhu J., et al. Emergence of Ceftazidime- and Avibactam-Resistant Klebsiella pneumoniae Carbapenemase-Producing Pseudomonas aeruginosa in China. Langelier CR, editor. mSystems. 2021 Dec 21;6(6):e00787–21.
[37] Weiser R, Green AE, Bull MJ, Cunningham-Oakes E, Jolley KA, Maiden MCJ, et al. Not all Pseudomonas aeruginosa are equal: strains from industrial sources possess uniquely large multireplicon genomes. Micro Genom 2019;5(7):e000276.
[38] Rosenberg C, Casse-Delbart F, Dusha I, David M, Boucher C. Megaplasmids in the plant-associated bacteria Rhizobium meliloti and Pseudomonas solanacearum. J Bacteriol 1982;150(1):402–6.
[39] Bryan LE, Semaka SD, Van den Elzen HM, Kinnear JE, Whitehouse RL. Characteristics of R931 and other Pseudomonas aeruginosa R factors. Antimicrob Agents Chemother 1973;3(5):625–37.
[40] Douarre PE, Mallet L, Radomski N, Felten A, Mistou MY. Analysis of COMPASS, a new comprehensive plasmid database revealed prevalence of multireplicon and extensive diversity of IncF plasmids. Front Microbiol 2020;11:483.
[41] Ravenhall M, Škunca N, Lassalle F, Dessimoz C. Inferring horizontal gene transfer. PLoS Comput Biol 2015;11(5):e1004095.
[42] Rocha EPC, Danchin A. Base composition bias might result from competition for metabolic resources. Trends Genet 2002;18(6):291–4.
[43] Sharp PM, Li WH. Codon usage in regulatory genes in Escherichia coli does not reflect selection for "rare" codons. Nucleic Acids Res 1986;14(19):7737–49. Oct 10.
[44] Suzuki H, Brown CJ, Forney LJ, Top EM. Comparison of correspondence analysis methods for synonymous codon usage in bacteria. DNA Res 2008;15(6):357–65.
[45] Wang Y, Yao L, Fan J, Zhao X, Zhang Q, Chen Y, et al. The codon usage bias analysis of free-living ciliates' macronuclear genomes and clustered regularly interspaced short palindromic repeats/Cas9 vector construction of stylonychia lemnae. Front Microbiol 2022;13:785889.
[46] Henry I, Sharp PM. Predicting gene expression level from codon usage bias. Mol Biol Evol 2007;24(1):10–2.
[47] Quax TEF, Claassens NJ, Söll D, van der Oost J. Codon bias as a means to fine-tune gene expression. Mol Cell 2015;59(2):149–61. Jul 16.
[48] Suzuki H, Saito R, Tomita M. A problem in multivariate analysis of codon usage data and a possible solution. FEBS Lett 2005;579(28):6499–504. Nov 21.
[49] Qh Z, Rq L, Yj W, Sl L. Identification of genes to differentiate closely related Salmonella lineages. PloS One [Internet] 2013 [cited 2024 Feb 1];8(2). Available from: https://pubmed.ncbi.nlm.nih.gov/23441160/.
[50] N. S, C. H Toward an efficient method of identifying core genes for evolutionary and functional microbial phylogenies. PloS one [Internet]. 2011 [cited 2024 Feb 1];6(9). Available from: ⟨https://pubmed.ncbi.nlm.nih.gov/21931822/⟩.
[51] Slater SC, Goldman BS, Goodner B, Setubal JC, Farrand SK, Nester EW, et al. Genome sequences of three agrobacterium biovars help elucidate the evolution of multichromosome genomes in bacteria. J Bacteriol 2009;191(8):2501–11.
[52] MacLean AM, Finan TM, Sadowsky MJ. Genomes of the symbiotic nitrogen-fixing bacteria of legumes. Plant Physiol 2007;144(2):615–22. Jun 1.