Incompatibility groups of *Pseudomonas* plasmids revisited: comprehensive analysis of R-factors and their replicons

AUTHORS[1]

Yosuke Nishimura[1+], Kensei Kaneko[2+], Tatsuya Kamijo[2+], Nanako Isogai[2], Maho Tokuda[2], Hui Xie[3], Yusuke Tsuda[4], Aki Hirabayashi[5], Ryota Moriuchi[6], Hideo Dohra[6,7,8], Kazuhide Kimbara[2], Chiho Suzuki-Minakuchi[3], Hideaki Nojiri[3], Haruo Suzuki[9]*, Masato Suzuki[5]* and Masaki Shintani[2,7,10]*

[1] Research Center for Bioscience and Nanoscience (CeBN), Japan Agency for Marine Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka-city, Kanagawa, 237-0061, Japan

[2] Department of Engineering, Graduate School of Integrated Science and Technology, Shizuoka University, 3-5-1 Johoku, Chuo-ku, Hamamatsu, Shizuoka, 432-8561, Japan

[3] Agro-Biotechnology Research Center (AgTECH), Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo, 113-8657, Japan

[4] Department of Clinical Laboratory Medicine, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8507, Japan

[5] Antimicrobial Resistance Research Center, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan

[6] Shizuoka Instrumental Analysis Center, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka, 422-8529, Japan

[7] Research Institute of Green Science and Technology, Shizuoka University, 3-5-1, Johoku, Chuo-ku, Hamamatsu, Shizuoka, 432-8561, Japan

[8] Department of Science, Graduate School of Integrated Science and Technology, Shizuoka University, 3-5-1, Johoku, Chuo -ku, Hamamatsu, Shizuoka, 432-8561, Japan

[9] Faculty of Environment and Information Studies, Keio University, 5322 Endo, Fujisawa, Kanagawa 252-0882, Japan

[10] Japan Collection of Microorganisms, RIKEN BioResource Research Center, 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

† Joint authors

* To whom correspondence should be addressed. Tel:+81-466-47-5099; Email: haruo@sfc.keio.ac.jp (Haruo Suzuki); Tel:+81-3-5285-1111; Email: suzuki-m@niid.go.jp (Masato Suzuki); Tel:+81-53-478-1181; Email: shintani.masaki@shizuoka.ac.jp (Masaki Shintani)

ABSTRACT

Bacterial plasmid incompatibility (Inc) groups, IncP-1 to IncP-14, have long been recognised as R-factors. These factors harbour antimicrobial resistance genes (ARGs) in *Pseudomonas* species (PIncs). Despite their importance, some PIncs plasmid remain underexplored at the sequence level. This study attempts to address this gap by determining the complete nucleotide sequences of several key plasmids. Using BLAST search and AlphaFold3-based protein structure prediction, we identified replication initiation proteins (RIPs) and origins of vegetative replication (*oriV*) for these plasmids. Notably, mini-plasmids containing these RIP and *oriV* sequences successfully replicated in *Pseudomonas aeruginosa*. Through pangenome analyses using the PLSDB database with the experimentally identified RIP gene reference, 2,351/59,895 plasmids were classified, and their core and accessory genes, including ARGs, were detected. Furthermore, phylogenetic analyses of the RIPs with publicly available protein sequences enabled the grouping of existing PIncs into distinct evolutionary lineages, providing a robust framework for identifying RIPs across a broader range of previously uncharacterised plasmids. The RIPs of 8,860 plasmids in the PLSDB were newly assigned. These findings offer crucial insights into the complex landscape of plasmid-mediated ARG dissemination, marking a significant advancement in the understanding of *Pseudomonas* plasmids across clinical and environmental contexts.

INTRODUCTION

Plasmids are self-replicating mobile genetic elements (MGEs) that play a key role in facilitating the exchange of genetic information among bacterial cells through conjugation or transformation. Plasmids, with their diverse functions, have emerged as the primary MGEs responsible for disseminating antimicrobial resistance genes (ARGs) among bacterial populations, which are classically known as R-factors and represent a threat to global public health and antimicrobial development (1). The accurate classification of plasmids is therefore indispensable for gaining valuable insights into the epidemiology of plasmid-mediated antimicrobial resistance, which would allow us to identify specific plasmids carrying ARGs. This precision of classification is paramount for conducting genomic epidemiological studies (2, 3). Historically, a primary method for classifying plasmids has been based on the concept of incompatibility (Inc) grouping, which relies on the phenotypic interactions between host cells harbouring two different plasmids (4). Plasmids that cannot co-reside within the same host cell line belong to the same Inc group, indicating that their replication or maintenance systems are similar.

ESKAPE pathogens, which include *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp., have acquired clinically significant ARGs via plasmids and are recognised as key multidrug-resistant bacterial species, and novel antimicrobial agents must be developed to combat these species (5). The genus *Pseudomonas*, including *P. aeruginosa*, has been extensively studied because of its diverse nature, widespread presence, and clinical significance as a causative agent of

various infections (6). Since the 1970s, a series of Inc groups (from IncP-1 to IncP-14) have been proposed based on incompatibility tests conducted on the genus *Pseudomonas* (Table 1) (7). In parallel, within the order *Enterobacterales*, which includes *Escherichia coli*, *Klebsiella* species, *Enterobacter* spp., and related bacteria, Inc groups ranging from IncA to IncZ were defined based on their host organisms (8). Notably, the same plasmid can be classified into different Inc groups depending on the bacterial host, with examples such as IncP-1 being equivalent to IncP, IncP-3 corresponding to either IncA or IncC, IncP-4 being the same as IncQ, and IncP-6 aligning with IncG. Based on these classifications, extensive studies have been conducted on the fundamental characteristics of these plasmids, including their replication, maintenance systems, and transmissibility (Table 1) (9–12).

However, the terminology surrounding these classifications can be confusing. The IncP-1 to IncP-14 groups, commonly referred to as "Inc groups in *Pseudomonas*," are often mistakenly labelled as “IncP” or "IncP groups". Notably, "IncP" specifically refers to a single group within the classification of plasmids in *Enterobacterales*, corresponding only to the IncP-1 group in *Pseudomonas*. To avoid confusion, we refer to the Inc groups in *Pseudomonas* as "PIncs." Recent advancements in sequence technologies have enabled the complete sequencing of several PIncs plasmids, providing detailed insights into their genetic makeup (13–16). Nevertheless, significant gaps exist in our understanding of certain PIncs groups, including IncP-5, IncP-11, IncP-12, IncP-13, or IncP-14, and only a partial DNA region is deposited for IncP-10 (Table 1). This is likely due to the narrow host range of some of these plasmids, which cannot be transmitted to *E. coli* from their original host *P. aeruginosa* via conjugation (17).

In this study, we present the complete nucleotide sequences of five R plasmids, i.e., Rms139 (IncP-2) (18), Rms163 (IncP-5) (19, 20), Rsu2 (IncP-9), RP1-1 (IncP-11) (21), R716 (IncP-12) (22), and pMG26 (IncP-13) (23, 24). Additionally, through *in silico* analyses, we identified a putative representative plasmid for IncP-10, similar to the archetype IncP-10 plasmid, R91-5 (25, 26). After identifying their replication-initiation protein (RIP) genes, we reclassified these plasmids in a publicly accessible database (27) and proposed the introduction of two new plasmid groups in *Pseudomonas*: IncP-15 and IncP-16. Furthermore, several other unclassified plasmids in the genus *Pseudomonas*, which harbour clinically important ARGs, such as genes conferring resistance to carbapenem and tigecycline, were examined. Two additional groups of plasmids, pQBR103-like and pSTY-like plasmids, were further discussed as unassigned PIncs groups. The elucidation of plasmid diversity and classification within PIncs is vital for deepening our understanding of genetic exchange mechanisms and the dissemination of ARGs. We aim to address the existing gaps in plasmid classification by proposing updated identification schemes of RIPs. Using RIP data elucidated in this study, we could accurately classify 2,351 plasmids within PLSDB, one of the plasmid databases containing 59,985 plasmids with their whole nucleotide sequences (27, 28), of which 27,126 have not yet been classified. The identified RIP information with previously known information was used to infer phylogenetic trees of putative RIPs from publicly available protein sequences (UniParc). Our analysis revealed the following: (i) except IncP-6, all other plasmids possess a winged-helix (WH) domain (named WH RIPs), exhibiting homology with each other, indicating a shared evolutionary origin. Based

116 on reconstructed phylogeny, we propose a bifurcated evolutionary scenario of WH RIPs; one
117 lineage that includes IncP-5, -9, -10, and -11, and another distinct lineage. (ii) IncP-6, in
118 contrast, belongs to the superfamily of archaeo-eukaryotic primases (AEPs) and represents a
119 distinct evolutionary lineage.

120 MATERIAL AND METHODS

121 Bacterial strains, plasmids, and culture conditions

122 The bacterial strains and plasmids used are listed in Table 2. *E. coli* and *Pseudomonas* strains
123 were cultured in Luria broth (LB) or M9 broth (29) at 30°C (*P. putida* KT2440) or 37°C (*P.*
124 *aeruginosa* PAO1 and its derivatives). LB plates containing 1.5% agar were used for filter
125 mating. Ampicillin (Ap, 50 µg/mL), piperacillin (Pip, 16 µg/mL), rifampicin (Rif, 50 µg/mL),
126 streptomycin (Sm 150 µg/mL), and tetracycline (Tc, 50 µg/mL for the others) were added to
127 the medium.

128 DNA manipulation

129 Total DNA from the bacterial strains was extracted using a NucleoSpin Tissue Kit (TAKARA BIO,
130 Japan). Polymerase chain reaction (PCR) was performed on a T100 thermal cycler (Bio-Rad
131 Laboratories, USA) using the KOD One PCR Master Mix (TOYOBO, Japan). The primers used in
132 this study are listed in Table S1. The following amplification conditions were used: 30 cycles at
133 98°C for 10 s, 67°C for 5 s, and 68°C for 1 s. The pUC-GW plasmid with DNA fragments
134 including putative RIP genes and *oriV*s of each plasmid, except those for RP1-1 and pMG26,
135 was synthesised by the custom service of Azenta Life Science, USA. Restriction enzymes (New
136 England Biolabs, USA or TAKARA BIO), the HiYieldTM Gel/PCR DNA fragments Extraction kit
137 (RBC Bioscience, Taiwan), NEBuilder Hifi DNA Assembly system (New England Biolabs), and
138 competent *E. coli* JM109 and DH5α cells (RBC Bioscience) were used for cloning the DNA
139 fragments. Plasmid DNA from the bacterial strains was extracted using the NucleoSpin Plasmid
140 EasyPure Kit (TAKARA BIO). Electroporation to introduce the plasmids into different bacterial
141 strains was performed using a MicroPulser electroporator (Bio-Rad Laboratories). All other
142 procedures were performed according to standard methods (29).

143 Conjugation assays

144 Filter mating assays were performed as described previously (30). In brief, donor and recipient
145 strains, *P. aeruginosa* PAO1808 with a plasmid and *P. aeruginosa* PAO1R or *P. putida* KT2440,
146 were cultured in LB with the appropriate antibiotic(s) (preculture). One millilitre of the above
147 culture in LB medium was harvested and suspended in fresh LB medium. Approximately 10^8
148 colony forming units/mL (CFU/mL) of the donor and recipient suspended in 150 µL LB were
149 mixed and then dropped on 0.2 µm pore-size filters (cellulose acetate, Advantec) on an LB agar
150 plate for 24 h at 37°C. The mixture on the filter was resuspended in PBS and then spread onto
151 LB plates with appropriate antimicrobials to select transconjugants (i.e., recipient cells with the
152 plasmid). Liquid mating assays were performed as follows. The precultured donor cells were

153 resuspended in 500 µL LB, mixed with 1 mL recipient cells for 2 h (200 rpm), and statically
154 incubated for 3 h at 37°C. The cells were then harvested, resuspended in PBS, and spread onto
155 LB plates with the appropriate antibiotics to select the transconjugants.

156 **Complete sequencing of the plasmids by next-generation sequencing**

157 The nucleotide sequences of the plasmid DNAs were determined as follows: whole genomic
158 DNA extracted from their host (*P. aeruginosa* PAO1808, a derivative strain of PAO1 (31)) was
159 sequenced using the HiSeq2500 platform (Illumina, USA). Trimmed high-quality short reads
160 (read length >140 bp and quality score >15) were assembled using SPAdes v3.14.0 using the
161 plasmid option (32). If circular contig(s) could not be found, the host chromosomal DNA was
162 removed by mapping the resultant contigs on the host genome sequences [their deposited
163 sequences, that is, DDBJ/GenBank accession number NC_002516 (*P. aeruginosa* PAO1)] using
164 Geneious Prime software v2020.01 (33). Subsequently, the plasmid reads were extracted from
165 the remaining contigs using SAMtools v1.7 (34) and SeqKit v0.8.0 (35) and reassembled using
166 SPAdes (36). For RP1-1 and pMG26, gaps in the plasmids were closed by PCR with the designed
167 primer sets based on the ends of contigs and Sanger sequencing of the PCR products.

168 **Gene annotation of plasmids**

169 Gene annotations of the plasmids were performed using DFAST-core v1.2.5 (37) with an in-
170 house database of plasmid sequences collected from NCBI RefSeq and then corrected
171 manually. Synteny analyses of plasmids were performed using Easyfig v2.2.2 (38). Other *in*
172 *silico* analyses were performed using Geneious Prime v2020.01 (33). The annotated plasmid
173 sequences were deposited under the following GenBank accession numbers: LC685025,
174 LC685026, LC685027, LC685593, and LC700336.

175 **Detection and phenotype identification of ARGs**

176 ARG sequences were detected using Staramr v0.10.0 and ResFinder database v2.2.1 updated
177 on 2023-10-27 (https://bitbucket.org/genomicepidemiology/resfinder_db) (39, 40) with the
178 following criteria: ≥ 90% identity and ≥ 60% coverage. Clinically important ARGs, including
179 genes conferring resistance to carbapenem, colistin, or tigecycline, in Gram-negative bacteria
180 were further detected based on the phenotype information of ARGs in AMRFinderPlus
181 v3.11.26 updated on 2023-11-16 (https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-
182 resistance/AMRFinder/) (41).

183 **Prediction of RIP genes and oriVs of each plasmid**

184 To identify the genes encoding the RIP and origin of vegetative replication (*oriV*) regions of
185 Rms139, Rms163, RP1-1, and R716, two prediction approaches were used; after coding
186 sequence (CDS) annotations using DFAST-core, protein.faa files generated by DFAST-core were
187 used for the following: (i) nucleotide sequence-based searches were performed using BLASTn;
188 (ii) amino acid sequence-based protein structure prediction and the structure-based searches
189 were performed as described later. The GC-skew value, which has been commonly used to

190 measure the asymmetry in the composition of two nucleotides, guanine (G) and cytosine (C),
191 along a DNA strand and is defined as (G-C)/(G+C), of the plasmids was used by Webskew
192 (https://genskew.csb.univie.ac.at/webskew) to detect the changes in the GC skew, which
193 indicates the DNA replication origins or termini (42, 43). Subsequently, the nucleotide or
194 amino acid sequences of putative CDSs around the above GC-skew shift point(s) were
195 subjected to methods (i) or (ii). The putative promoter region of the RIP gene was predicted
196 using BPROM
197 (http://www.softberry.com/berry.phtml?topic=bprom&group=help&subgroup=gfindb) (44).
198 Subsequently, the predicted RIP and *oriV* sequences of each PIncs were cloned into the pUC-
199 GW-Amp vector (Azenta Life Science). Transformation of *P. aeruginosa* PAO1808 was
200 performed with each of the resultant plasmids (at most 472 ng/µL), and the cultures were
201 spread onto LB+Pip plates. After incubation at 37ºC overnight, colonies on the plates were
202 isolated. Next, the presence of the mini-plasmid was confirmed by PCR with specific primers
203 after extraction of total DNA of transformants.

204 **Protein structure prediction of RIPs**

205 AlphaFold3-mediated protein structure prediction was performed using the amino acid
206 sequences of the RIP gene products in representative *Pseudomonas* plasmids
207 (https://figshare.com/s/9554d4fc982d7929fb77) on the AlphaFold server (45). A structure-
208 based protein homology search using the predicted structural models was performed for the
209 PDB100 database on the FoldSeek server in 3Di/AA mode (46). The predicted and
210 experimental structures were visualised and compared using PyMOL v3.0.3. The predicted
211 structures were coloured based on the root mean square deviation (RMSD) from the reference
212 experimental structures, i.e., the Pi or Rep19A proteins, which are RIPs of the R6K and pCA-347
213 plasmids, respectively (47, 48), using the ColorByRMSD script
214 (https://pymolwiki.org/index.php/ColorByRMSD). The distances between aligned C-alpha atom
215 pairs were stored as B-factors of these residues, which are coloured using a colour spectrum,
216 with blue specifying the minimum pairwise RMSD and red indicating the maximum. Unaligned
217 residues were coloured grey.

218 **Identification, retrieval, and re-annotation of plasmids from public databases**

219 Complete plasmid sequences were downloaded in FASTA format using the Entrez
220 Programming Utilities (E-utilities; https://www.ncbi.nlm.nih.gov/books/NBK25501/) from a
221 public database (PLSDB 2023_11_03_v2) (https://figshare.com/s/915a336484ee084ffbea) (27,
222 28). To classify the plasmids in the PLSDB based on their rep sequence similarity, the RIP gene
223 sequences were detected using Staramr v0.10.0 (https://github.com/phac-nml/staramr) and
224 rep databases of MOB-suite v3.1.8 updated on 2023-12-12 (https://github.com/phac-
225 nml/mob-suite) (https://figshare.com/s/233faa319102390218e5) (49), PlasmidFinder v2.1
226 updated on 2023-12-04 (https://bitbucket.org/genomicepidemiology/plasmidfinder_db)
227 (https://figshare.com/s/7de82d807b4a1630f047) (50), and our custom-made *Pseudomonas*
228 plasmid RIP gene database (termed repP database,
229 https://figshare.com/s/28e3204f6991d6ce0c11) with the following criteria: ≥90% identity

230 and ≥60% coverage. The repP database was constructed using experimentally identified RIP
231 genes from the 87 reference *Pseudomonas* plasmids shown in Table S2. Furthermore,
232 duplicated RIP sequences between the MOB-suite and repP databases (Table S3) and incorrect
233 RIP sequences, including uncertain IncP and rep_cluster_1115, were manually removed from
234 the MOB-suite database. The modified MOB-suite database
235 (https://figshare.com/s/3e3ced8633136259f678) was used to detect multi-replicon
236 *Pseudomonas* plasmids, which are single plasmids that can belong to more than one group
237 other than the PIncs groups. Among the candidate IncP-1 plasmid sequences based on the
238 similarity to the RIP sequences, artificially constructed sequences not explicitly labelled as such
239 ("other entries;other sequences;artificial sequences") in the NCBI taxonomy database
240 (https://www.ncbi.nlm.nih.gov/taxonomy) were also present. We identified six IncP-1
241 plasmids (PD584_L3, PD689_delTT1, pJM105, pJM50, pRK2013, and pUZ8002) through a
242 literature survey, and manually removed these sequences from the putative IncP-1 plasmid
243 sequences.

244 For the final dataset, the sequence length (bp) and GC content (%) were calculated using
245 SeqKit v2.3.0 (https://bioinf.shenwei.me/seqkit/) (35). Plasmid sequences belonging to each
246 PIncs group were re-annotated using DFAST-core, with a reference genome from each group
247 to ensure consistency in the functional annotations. GFF3 annotation files produced by DFAST-
248 core were subsequently used as input files for PIRATE, a toolbox for pangenome analysis (51).

249 **Pangenome analysis of plasmids**

250 A pangenome, i.e., the entire set of protein genes from all members of a defined group (all the
251 plasmids of the PIncs group), was analysed using PIRATE
252 (https://github.com/SionBayliss/PIRATE). Here, we used DIAMOND v0.9.14 (52) rather than
253 BLAST (53) to identify homologous genes (protein families) in PIRATE. PIRATE was run on
254 default parameters with a high-scoring segment pair (HSP) query length threshold of 0.5, a
255 sequence similarity step range of 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, and 98%, and an MCL
256 v14.137 inflation value of 2. A locally optimal alignment with no gaps is called a HSP in BLAST
257 (54). MCL is a Markov cluster algorithm (https://micans.org/mcl/). The core genes in each
258 PIncs group were those that were present in more than 95% of the plasmids included in the
259 pangenome analysis. Statistical computing and result visualisation were performed using R
260 v4.1.3 (https://www.r-project.org/). The output of SeqKit and PIRATE are available in Figshare
261 (https://figshare.com/s/0f2aa75d1eb4c5848a32).

262 **Phylogenetic analysis of the RIPs**

263 Detailed methods are provided in Supplementary Text S1 and illustrated in Figure S1. In brief,
264 to analyse the evolutionary relationships of the RIPs of PIncs, we identified 2,374 RIPs that
265 originated from 2,332 PInc plasmids. Subsequently, 251 non-redundant proteins at 100%
266 identity were extracted using MMseqs cluster v15-6f452 (55). Sequence homology between
267 the non-redundant proteins was assessed using BLASTp (53), and the resultant hits were
268 visualised via CPAP (https://github.com/yosuken/CPAP). To detect remote homology between

269 PIncs, we extended the sequence diversity by collecting homologous sequences through a two-
270 step homology search process. In the first step, PLSDB proteins predicted with orfM (n =
271 20,331,985) (56) were searched using hmmsearch of HMMER v3.3.2 (57). In the second step,
272 UniParc proteins (downloaded on Feb 2024; n = 607,912,929) were searched using
273 hmmsearch. We obtained 6,948 non-redundant sequences (at 60% identity) that were
274 classified into 11 homologous groups. To investigate whether these groups exhibit homology
275 with each other, a highly sensitive hidden Markov model (HMM)–HMM comparison was
276 performed using HHblits v3.3.0 (58). We extracted the homologous regions of the alignments
277 of the 10 groups (containing 6,306 sequences in total). The secondary structures of the
278 homologous regions were predicted and visualised based on the crystal structures (PDB: 2NRA,
279 4PT7) and alphafold3 predictions using iCn3D
280 (https://www.ncbi.nlm.nih.gov/Structure/icn3d/) (59). After removing the gappy sequences,
281 we realigned the remaining 6,265 sequences using mafft v7.455 (60).

282 A maximum-likelihood phylogeny of WH RIPs was reconstructed using the alignment and IQ-
283 TREE v2.2.2.6 (61). The best fit model (Q.pfam+F+R4) was selected using ModelFinder (62). We
284 evaluated the supporting values of the best tree with UFBoot2 (63) via IQ-TREE and transfer
285 bootstrap expectation (TBE) via BOOSTER v0.1.2 (64). We discarded the abnormally long
286 branches using treeshrink (65) and redundant sequences. The resultant tree was rooted based
287 on the PG5/PG9/PG10 superclade (n = 400 sequences) using gotree v0.4.0 (66). The tree-
288 associated UniParc proteins and PLSDB plasmids were identified based on sequence homology
289 search results, and subsequently, the consensus taxonomic affiliation for each leaf was
290 identified. For a detailed description, see Supplementary Text S1 and Figure S2. Phylogenetic
291 tree–related files are available in Figshare (https://figshare.com/s/7303f34f6d8b12214f01).

292 RESULTS AND DISCUSSION

293 Overview of plasmid sequences

294 The genomes of *P. aeruginosa* PAO1808, each harbouring one of the six plasmids (Rms139,
295 Rms163, Rsu2, RP1-1, R716, and pMG26), were sequenced. Thus, the complete nucleotide
296 sequences for each R factor were successfully obtained (Table 1). Notably, although previous
297 studies have mostly focussed on IncP-9 plasmids, we also determined the complete nucleotide
298 sequence of an ancestral R plasmid from the IncP-9 group, known as Rsu2 or R2 in certain
299 literature, which was originally isolated at Shinshu University in Japan (67, 68). Regarding
300 pMG26, no circular contigs were identified following the removal of reads mapped to the
301 reference genome, the chromosome of *P. aeruginosa* PAO1. Instead, a unique 99-kb DNA
302 region was identified within the chromosome of its host, PAO1808, which distinguished it from
303 other plasmid hosts.

304 Identification of the RIP gene in each R plasmid

305 RIPs of the representative plasmids in each PInc group were identified by nucleotide sequence-
306 based search and/or structure-based search (see Materials and Methods).

307 The IncP-2 plasmid Rms139 harbours a single RIP gene, called *repP-2A*, with its putative *oriV*
308 identified between nucleotides 2,395 and 3,150 (23). This *oriV* region contains nine repeats of
309 an 11-bp sequence [5'-TA(G/C)TCAGGTTT-3'], two putative DnaA boxes [5'-
310 TT(A/T)(T/G)C(T/C)(A/T)CA-3'], and an A+T-rich region (Figure 1). A mini-Rms139A construct,
311 containing the *repP-2A* gene and *oriV*, was successfully introduced into *P. aeruginosa* via
312 electroporation, achieving a transformation frequency of 5.0×10^3 CFU/µg-DNA. No
313 transformants were detected with the control vector pUC-GW-AMP ($<3.4 \times 10^0$ CFU/µg-DNA).

314 Initially, no RIP gene was detected in the IncP-5 plasmid Rms163 through nucleotide sequence-
315 based searches. Subsequently, a GC-skew plot of Rms163 was generated (Figure S3), and three
316 nearby CDSs were selected for three-dimensional structure prediction. One CDS showed
317 structural similarity to the replication initiator A family proteins, specifically to Rep19A of pCA-
318 347 with RMSD = 4.2 (Figure 2B). Downstream of this gene, a putative *oriV* region was
319 identified with characteristic repeats, palindromic sequences, a putative DnaA box, and an
320 A+T-rich region (Figure 1). Transformation of *P. aeruginosa* PAO1 with a mini-Rms163A
321 construct containing the RIP gene candidate and *oriV* resulted in a transformation efficiency of
322 4.4×10^1 CFU/µg-DNA. This candidate gene was subsequently named *repP-5A*.

323 Although the complete sequence of the plasmid previously identified as IncP-10 is not
324 available in public databases, a 2.7-kb DNA sequence from the archetype IncP-10 plasmid R91-
325 5 (25) was used as the query. This sequence was previously identified as the replication region
326 of R91-5 through incompatibility testing (26). We identified 45 plasmids in the PLSDB with
327 regions similar to this 2.7-kb fragment, including an unnamed plasmid from the *P. aeruginosa*
328 strain PAB546 (GenBank accession number NZ_MN433456.1), which we tentatively named
329 pPAB546. While no RIP gene was initially annotated in this plasmid, a manual search similar to
330 that used for IncP-5 led to the identification of a putative RIP similar to Rep19A (RMSD = 5.4)
331 (Figure 2B), encoded by a gene near the conserved 2.7-kb region (Figure 1). The
332 transformation of mini-pPAB546 into *P. aeruginosa* yielded a frequency of 1.7×10^2 CFU/µg-
333 DNA.

334 The IncP-11 plasmid RP1-1 possessed a putative CDS classified as rep_cluster_2025 by the
335 MOB-suite package (49), located at nucleotides 54,711-54,953 of the RP1-1 sequence
336 (GenBank accession number LC700336). A nearby *oriV* region featured four 12-bp repeats and
337 an A+T-rich region, although no DnaA box was identified (Figure 1). Initial transformation
338 attempts with a pUC-GW-AMP construct harbouring these regions failed to yield
339 transformants ($<8.0 \times 10^1$ CFU/µg-DNA). However, a downstream CDS, showing 61% identity
340 with the replication initiator protein RepA from a *P. stutzeri* plasmid-like contig, was cloned
341 along with the *oriV* region. The constructed mini-RP1-1A successfully transformed *P.*
342 *aeruginosa* at a frequency of 1.3×10^2 CFU/µg-DNA (Figure 1), leading to the designation of
343 the CDS as *repP-11A*.

344 Similar to the IncP-2 plasmid, the IncP-12 plasmid R716 contained a putative RIP gene, *repP-*
345 *12A*, encoding a trans-acting replication function (TrfA) family protein (pfam07042), which is
346 analogous to the RIPs found in IncP-1 plasmids. Downstream of *repP-12A*, we identified eight
347 repeats of a 14-bp sequence (5'-CGTCGTCCGACGCA-3'), one putative DnaA box (5'-
348 TTATCCATA-3'), and an A+T-rich region (Figure 1). A mini-R716A construct containing *repP-12A*
349 and *oriV* was successfully introduced into *P. aeruginosa* via electroporation, with a
350 transformation frequency of 1.1×10^3 CFU/µg-DNA.

351 **pMG26 (IncP-13) is not a plasmid but an integrative and conjugative element (ICE)**

352 The plasmid pMG26, previously identified as a multidrug resistance plasmid and classified as
353 IncP-13 (17), did not yield a circular contig during assembly, even after removing reads
354 mapping to the PAO1808 chromosome. Instead, a 99-kb DNA region was detected within the
355 chromosome of its host, PAO1808, which was not conserved in the hosts of other plasmids.
356 This observation led to the hypothesis that pMG26 might function as an ICE, similar to the
357 reclassification of FP8 as ICEPae1161 (69).

358 To investigate this hypothesis, PCR analyses were performed to determine whether an
359 intermediate circular DNA form of the 99-kb DNA region could be generated in the host
360 PAO1808. As shown in Figure S4, PCR products were obtained using three different primer
361 sets, suggesting that the pMG26 could indeed form a circular intermediate within its host.
362 Notably, 49-bp direct repeats (5'-
363 ATGGTGGGTCGTGTAGGATTCGAACCTACGACCAATTGGTTAAAAGCCA-3') were identified at both
364 ends of the integrated DNA region. The predicted circular form of pMG26 was 99,171 bp in
365 length and was integrated into the $tRNA_{Lys}$ locus situated between the *lepA* and *clpB* region in
366 the *P. aeruginosa* PAO1 chromosome (5,086,973 nt, NC_002516).

367 The 99-kb DNA region of pMG26 encoded a MOB_H type relaxase and MPF_G family type IV
368 secretion system (T4SS) (Table 1). To test its conjugative transferability, both filter mating and
369 liquid mating assays were performed between *P. aeruginosa* PAO1808 (donor) carrying pMG26
370 and *P. putida* KT2440 (recipient). The element was successfully transferred from the donor to
371 the recipient strain, with transfer frequencies of $1.4\text{-}1.9 \times 10^{-5}$ per donor (filter mating) and
372 $0.28\text{-}3.4 \times 10^{-5}$ (liquid mating) per donor (Table 3). These results strongly suggest that pMG26
373 functions as an ICE.

374 Additionally, PCR analysis revealed that pMG26 integrated at three different $tRNA_{Lys}$ sites on
375 the KT2440 chromosome (PP_t09, PP_t11 and PP_t13). Notably, the entire DNA region of
376 pMG26 exhibited a high degree of similarity with that of pKLC102 (70) (Figure S5), a member of
377 the PAPI-1 family of ICEs, which are abundant in *P. aeruginosa* (70). Further analyses using
378 MOB-suite indicated that rep_cluster_322 was present in both pMG26 and pKLC102. The
379 putative *oriV* region of pKLC102 (70) was also identified in pMG26.

380 Despite these similarities, no transformants were detected when a mini-pMG26A containing
381 the rep_cluster_322 CDS and putative *oriV* region of pMG26 was introduced into *P. aeruginosa*
382 (below the detection limit, <2.0 x 10^1 CFU/µg-DNA). This result was not unexpected because
383 the autonomous replication of ICE is typically mediated by rolling-circle replication involving
384 their relaxases (71).

385 These findings suggest that pMG26, originally classified as an IncP-13 plasmid, functions as an
386 ICE, with a potential role in the horizontal transfer of multidrug resistance genes in
387 *Pseudomonas* spp. The high similarity with pKLC102 further supports this reclassification and
388 underscores the significance of ICEs in the evolution and dissemination of resistance
389 determinants.

390 **Transferability of other R-factors**

391 The transferability of R plasmids Rms139 (IncP-2), Rms169 (IncP-5), Rsu2 (IncP-9), RP1-1 (IncP-
392 11), and R716 (IncP-12) was assessed using both filter mating and liquid mating assays. All
393 plasmids were successfully transferred from *P. aeruginosa* PAO1808 to *P. aeruginosa* PAO1 via
394 filter mating on solid media, indicating their self-transmissible nature (Table 3). Additionally,
395 Rms139 was successfully transferred to *P. putida* KT2440 (Table 3).

396 However, no transconjugants of Rms163 were detected in the liquid mating assay (below the
397 detection limit, <2.0 × 10^{-7} per donor), suggesting that the conjugation system of Rms163 is
398 specifically adjusted for transfer under solid conditions. This may reflect the differences in the
399 environmental conditions or surface requirements for effective mating formation (72, 73).

400 MOB-suite analysis revealed that the IncP-2 plasmid Rms139 carried the MPF_I family T4SS and
401 MOB_P type relaxase gene (Table 1). In contrast, neither the previously known MOB type
402 relaxase nor the MPF family T4SS components were identified in the IncP-5 plasmid Rms163 or
403 IncP-12 plasmid R716. Instead, both Rms163 and R716 were found to harbour defect in
404 organelle trafficking (Dot) or intracellular multiplication (Icm) systems, which are known as
405 type IVB secretion systems (74). Similarly, the IncP-11 plasmid RP1-1 (GenBank accession
406 number LC700336) lacked the previously known MOB type relaxase and MPF families but
407 possessed putative genes for relaxase (34,618-35,376 nt), type IV coupling protein (32,822-
408 30,702 nt), and proteins involved in mating pair formation.

409 These findings confirm that all the R-factors sequenced in this study are self-transmissible,
410 although they utilise diverse mechanisms for conjugation. This diversity in conjugation systems
411 highlights the varied strategies employed by plasmids to spread across bacterial populations,
412 which has important implications for the dissemination of antibiotic resistance genes.

413 **Reclassifications of plasmids in PLSDB based on RIPs of PIncs**

414 Based on the nucleotide sequences of the identified RIP and *oriV* regions of the plasmids of
415 IncP-1 to IncP-13, new Inc groups within the *Pseudomonas* genus could be proposed. In our
416 previous studies, novel plasmid groups hosted by *Pseudomonas*, known as PromA group and
417 pSN1216-29-like plasmids, were identified (75–78). As the nucleotide sequences of their RIP
418 gene and *oriV* region differed from those of IncP-1 to IncP-13, these new plasmid groups were
419 designated as IncP-15 and IncP-16, respectively.

420 Several other important *Pseudomonas*-derived plasmids carrying ARGs have not yet been
421 assigned to any Inc group. For instance, pQBR103 is a large 425-kb plasmid identified in
422 *Pseudomona*s strains colonising the leaf and root surfaces of sugar beet plants (79, 80). This
423 plasmid contains two putative RIP genes located at 339,909-341,066 nt and 388,322-389,491
424 nt. Notably, the latter RIP gene exhibited low identity (46.2% at the nucleotide level) with
425 *repP-2A*. Recently, similar large-size plasmids (>400 kb) have been identified, which carry the
426 *tmexCD-toprJ* gene cluster, encoding a resistance-modulation-division (RND) family efflux
427 pump (81), conferring resistance to multiple classes of antimicrobials including tigecycline.

428 Another plasmid of interest is pSTY, which was identified in the styrene degrading
429 *Pseudomonas* strain VLB120 (82). Recently, plasmids similar to pSTY have been reported to
430 carry carbapenem resistance genes, bla_{VIM} (83). Their RIP genes and *oriV* regions were
431 predicted, leading to their provisional classification as Inc_{pSTY}, despite the absence of
432 incompatibility testing (83). The identification and classification of such plasmids are crucial
433 because they expand our understanding of the diversity and evolutionary trajectories of ARG-
434 bearing plasmids in *Pseudomonas*.

435 To identify and accurately classify similar plasmids within the publicly available PLSDB
436 (2023_11_03_v2, containing 59,895 plasmids) into specific PIncs, we prepared custom-made
437 RIP reference sequences (Table S2). These reference sequences included the newly identified
438 RIP gene sequences of PIncs, excluding IncP-8 and IncP-13 (ICEs) and IncP-14 (not available at
439 the time). Using these custom sequences, 2,357 plasmids were identified that could be
440 classified into PIncs (Table 1, S4). This approach allowed for the classification of 530 plasmids
441 that were not previously classified by the PlasmidFinder database and 1,514 plasmids that
442 were not previously classified by the MOB-suite package database, resulting in a more
443 comprehensive plasmid classification. Notably, our custom-made RIP reference corrected 109
444 misclassifications identified by the MOB-suite package database, underscoring the accuracy of
445 our approach (https://figshare.com/s/915a336484ee084ffbea).

446 The distribution of the 2,357 plasmids across each PIncs group is summarised in Table 1 and
447 Table S4. PIRATE analyses were performed to identify core genes present in 95–100% of the
448 plasmids within each PIncs group (Table 1, S5-1~S5-14), revealing several key features. For
449 instance, the genome sizes of IncP-2, pQBR103-like plasmids, and pSTY-like plasmids were
450 exceptionally large (Table 1, S4, Figure 3A). Additionally, the IncP-1 and IncP-4 plasmids
451 exhibited a wider range of G+C content compared to other groups of plasmids (Figure 3B).
452 Regarding the numbers of core genes, IncP-2, IncP-12, and pQBR103-like plasmids had more
453 core genes than other groups (Table 1, S4, Figure 3C). Furthermore, over 20% of the plasmids

454 in the IncP-2, P-3, P-4, P-6 P-10, and P-11 groups were identified as multi-replicons, meaning
455 that they carried additional RIP genes besides the *repA* gene of PIncs (Table 1, S4, Figure 3D).
456 Importantly, more than 60% of the IncP-2, P-3, P-4, P-6, P-10, and pQBR103-like plasmids
457 harboured ARGs, with several of these groups, IncP-2, P-6, P-10, and pQBR103-like plasmids,
458 containing ARGs of significant clinical importance (Figure 3D).

459 **IncP-1 (=IncP) plasmids**

460 In our previous study, 14 subgroups within IncP-1, denoted as IncP-1αβγδεζηθικλμορ, were
461 proposed based on the incompatibility tests (78). In this study, we tentatively propose an
462 additional subgroup, IncP/P-1π, which includes 41 plasmids (Table S4), including pEP5289.
463 Although pEP5289 was previously identified as a member of the IncP-1 group found in the
464 genus *Neisseria* (84), it has not been experimentally demonstrated whether this plasmid is
465 incompatible with other IncP-1 plasmids. Furthermore, the host range of these plasmids is
466 limited to the genus *Neisseria*, a feature that may distinguish them from other IncP-1 plasmids.
467 A total of 25 core genes were identified across 321 plasmids in this group, the average
468 numbers of which were 65.7 (Table 1, S4, S5-1).

469 **IncP-2 plasmids**

470 Plasmids belonging to the IncP-2 group exhibit significantly larger average DNA sizes of
471 approximately 440 kb. Notably, 293 core genes were identified within this group (Table 1, S4,
472 S5-2), a number significantly higher than that in the other PIncs group. This was around 53% of
473 the average gene numbers (293/549.8) of 87 plasmids in IncP-2 (Table S4). As previously
474 reported (85), the megaplasmid family known as pBT2436 and pRBL16 (14, 15), associated
475 with ARGs, has been recognised as part of the IncP-2 plasmids. Given their large size, specific
476 gene sets in the plasmids are likely crucial for maintaining these plasmids within their hosts.
477 Interestingly, one core gene product exhibited identity with the Tus protein (Table S5-2),
478 known for binding to DNA replication terminus (*ter*) sites of replicons (86). This protein,
479 encoded by rep_cluster_1115 in the MOB-suite database, was found not only in all IncP-2
480 plasmids but also in nine other large-sized plasmids (113–527 kb) and, therefore, was removed
481 from our custom-made RIP reference sequences
482 (https://figshare.com/s/3e3ced8633136259f678).

483 **IncP-3A (=IncA) or IncP-3C (IncC)**

484 The IncP-3 group, also known as the IncA/C group, was recently reclassified into two separate
485 Inc groups, IncA and IncC (87). Although RIP genes within IncA and IncC plasmids are highly
486 conserved (>98% identity), the differences in the iteron numbers, spacing, and locations could
487 make them compatible (87). PIRATE analysis revealed that these plasmids share 48 core genes
488 (24.0% of average gene numbers, 48/200.1) (Table 1, S4, S5-3), but the accessory genes in the
489 IncP-3 plasmids exhibited greater diversity (Figure 3C). Notably, most IncP-3 (94.5%) plasmids

490 carried at least one ARG (Table 1, S4), indicating that these plasmids are significant vectors for
491 ARGs.

492 **IncP-4 (=IncQ) plasmids**

493 IncP-4 plasmids, also known as the IncQ group, exhibit a wide range of DNA sizes, from 1.7 to
494 412 kb, with an average of 121 kb (Table 1, Figure 3A). These plasmids are characterised by the
495 presence of *repBAC* genes, which are essential for their single-strand displacement replication
496 (12). However, only two core genes *repA* and *repC* (*repP-4A*) were consistently observed across
497 this group, which was a rather small ratio for their average gene numbers, i.e., 1.5% of 134.1
498 (Table 1, S4, S5-4). This was possibly due to the presence of multi-replicons in the IncP-4
499 group, with only 29.1% plasmids containing a single replicon (Table S4).

500 **IncP-5 plasmids**

501 Only one plasmid closely related to Rms163, an IncP-5 plasmid, was identified in the PLSDB
502 (Table 1). This limited representation in the database may not reflect the true prevalence of
503 these plasmids in natural environments. ARGs in Rms163 were located within a Tn*3*-like
504 transposon and a remnant of class 1 integron (Figure S6), with nucleotide sequences showing
505 100% identity to segments of Tn*5393* and class 1 integron previously observed in an IncP-9
506 plasmid, pSN0726-36, isolated from a soil sample in Japan (78). Interestingly, the *repP-5A* gene
507 in Rms163 exhibited 100% identity with the corresponding gene in pPALA50 (GenBank
508 accession number CP111035.1), a gene recently discovered in *P. aeruginosa* isolated from the
509 sputum of a patient with cystic fibrosis (88). Despite the absence of ARGS in pPALA50, the high
510 conservation of other genetic elements between Rms163 and pPALA50 is notable (Figure S6).

511 **IncP-6 (=IncG) plasmids**

512 IncP-6 plasmids, similar to IncP-4, exhibited diverse DNA sizes ranging from 12 to 497 kb
513 (Figure 3A). Three core genes were identified, including *repP-6A*, *parC* (involved in plasmid
514 partitioning (89)), and a putative recombinase/transposase gene (Table S5-5). This was a small
515 ratio of average gene numbers, i.e., 5.7% (3/52.7) (Table S4). The archetype plasmid of IncP-6,
516 Rms149, is incompatible with IncU plasmids including pRA3 and pFBAOT6 (90). Although they
517 share the *repP-6A* and *parC* genes, the backbone genes between IncP-6 and IncU plasmids
518 differ (90, 91), suggesting the potential need for further subgroup classification. In this study,
519 the IncP-6 (IncG) and IncU plasmids were separately recognised.

520 **IncP-7 plasmids**

521 IncP-7 plasmids shared 10 core genes (5.2% per 191.9 average gene numbers, Table S4),
522 including *repA* (*repP-7A*), and *parWAB* genes, which are involved in plasmid partitioning (92)

523 (Table S5-6). Interestingly, some IncP-7 plasmids, including pWW53 and pND6-1 (93, 94),
524 lacked genes for conjugative transfer, likely explaining why the numbers of core genes were
525 small (Table S4). Among these 23 plasmids, only four carried ARGs, whereas nine others
526 carried catabolic genes, including pCAR1(note that three derivative plasmids has been
527 deposited), pDK1, pWW53, pND6-1, pPHE24 (95), and unnamed plasmids (GenBank accession
528 number NZ_CP071901) (96).

529 **IncP-9 plasmids**

530 The 10 core genes of IncP-9 plasmids (9.3% per average core gene numbers, 10/107.6 (Table
531 S4)) included *rep* (*repP-9A*), genes for plasmid partition, and a single-strand DNA binding
532 protein (*ssb*) (Table S5-7). The plasmid sizes ranged from 38 to 483 kb. Notably, some smaller
533 plasmids did not carry any genes for conjugative transfer (Table S5-7), whereas the largest
534 plasmid, pKF715A (GenBank accession number NZ_AP015030.1), carried the *bph-sal* element
535 (97). IncP-9 plasmids, which have been well-studied in *Pseudomonas*, show diversity based on
536 the partial sequences of *rep* and *oriV*, leading to proposals for several subgroups (98).
537 Although many IncP-9 plasmids carry ARGs, some, such as pWW0 and NAH7, are catabolic
538 plasmids with lower sequence identity in *repP-9A* and other core genes compared to ARG-
539 carrying IncP-9 plasmids (data not shown).

540 **IncP-10 plasmids**

541 All the identified IncP-10 plasmids were hosted by *P. aeruginosa*. More than 50% of them
542 contained ARGs for carbapenem (Table 1). Among the 49 plasmids in the IncP-10 group, the
543 sequence lengths ranged from 29,402 to 106,679 base pairs (bp). Pangenome analysis using
544 PIRATE identified 209 gene families across these plasmids, with 19 core genes present in 95–
545 100% of the plasmids (29.0% per average gene numbers, 19/65.6) (Table S4), including
546 putative genes for plasmid replication and maintenance (Table S5-8).

547 **IncP-11 plasmids**

548 IncP-11 plasmids, which are also hosted by *P. aeruginosa* (99), shared 19 core genes, including
549 *repP-11A*, plasmid partition genes, and genes related to conjugative transfer (Table S5-9). The
550 ratio of the core genes per average gene numbers of IncP-10 plasmids was 47.6% (19/65.6,
551 Table S4). Several plasmids were observed in clinical isolates, indicating that IncP-11 plasmids
552 are significant carriers of ARGs in human pathogens. Importantly, pOXA-198 (GenBank
553 accession number MG958650), previously proposed as an IncP-11 plasmid (100), lacks any
554 genes resembling the *repP-11A*, suggesting that it may not belong to the traditionally
555 recognised IncP-11 group (101).

556 **IncP-12 plasmids**

557 Nine plasmids hosted by *P. aeruginosa* were categorised as IncP-12 plasmids (Table 1), with
558 165 core genes, which was a rather large ratio, 74.2% per their average gene numbers
559 (165/222.4, Table S4), although several core genes encoded putative hypothetical proteins
560 (Table S5-10). This indicated that nine plasmids in IncP-12 had high genetic conservation,
561 although none of them other than R716 carried any known ARGs.

562 **IncP-15 (PromA)**

563 A total of 42 plasmids were identified as IncP-15/PromA plasmids, which share 30 core genes,
564 which was 50.3% of their average gene numbers (30/59.6, Table S4). These core genes
565 included genes involved in replication and conjugative transfer (Table S5-11). Interestingly,
566 several of these plasmids do not carry any known accessory genes and exhibit very conserved
567 genetic structures in different subgroups (75, 77).

568 **IncP-16 (pSN1216-29) plasmids**

569 A total of 24 plasmids were classified as IncP-16 (pSN1216-29), with their sizes ranging from 35
570 to 174 kb. These plasmids shared 17 core genes (28.9% of their average gene numbers,
571 17/58.9) (Table S4) including *repA* and some genes involved in conjugative transfer (Table S5-
572 12). Some of these plasmids, such as pCTX-M15 and pSRC119-A/C, are multi-replicons, which
573 may explain the reduced number of shared genes. Three IncP-16 plasmids carried ARGs.

574 **pQBR103-like plasmids**

575 Similar to IncP-2, pQBR103-like plasmids have larger average DNA sizes (~515 kb), with 229
576 core genes identified (Table S5-13). These plasmids are known as pJBCL41-type plasmids and
577 important as they are associated with antimicrobial resistance, especially clinically relevant
578 carbapenem and tigecycline resistance (102, 103).

579 **pSTY-like plasmids**

580 pSTY-like plasmids exhibited significant diversity in their genetic content. Pangenome analysis
581 using PIRATE identified 1,073 gene families across 15 plasmids in this group, with 66 core
582 genes (Table S5-14). Some of these plasmids have been linked to clinically relevant ARGs, such
583 as carbapenem and tigecycline resistance (83, 103), making them important targets for further
584 research.

585 **Phylogenetic analysis of RIPs**

586 To investigate the evolutionary relationships of 15 PIncs, we performed all-against-all BLASTp
587 and analysed the homologous relationships between 251 non-redundant RIP representatives

588 (at 100% sequence identity) (Figure 4). The degree of sequence similarity of RIPs within each
589 PInc was found to vary depending on the PInc. Some PIncs show a high degree of sequence
590 variation, namely IncP-1 (34.05% identity for the most divergent pair), IncP-4 (28.81%), and
591 IncP-15 (56.14%). In contrast, the other PIncs contain less diverse RIP proteins (>73% identity
592 within each PInc) (Table S6). We rarely detected homologous regions with BLASTp between
593 the most sequence pairs of different PIncs, except for some specific combinations of PIncs,
594 such as IncP-2/IncP-3/pQBR103, IncP-1/IncP-12, and IncP-10/IncP-11, which exhibited
595 sequence similarity between different PIncs (Figure 4).

596 The low sequence diversity of the RIPs is not suitable for the detection of remote homology
597 between the RIPs of 15 PIncs. Thus, we extensively collected homologous sequences of RIPs
598 through a two-step homology search process (Figure S2A). In the first step, we collected
599 homologous proteins of the RIPs from the entire PLSDB plasmid dataset. We obtained 1,885
600 non-redundant RIPs with 90% identity, which were reclassified into 12 groups of homologous
601 sequences (i.e., some PIncs were merged including RIPs of similar sequences; see Figure S2). In
602 the second step, we collected homologous sequences of RIPs of the first step, from the
603 comprehensive protein database UniParc, and obtained 6,948 non-redundant RIPs with 60%
604 identity (87,818 sequences in total), consisting of 11 groups of homologous sequences (Figure
605 S2).

606 We then analysed the homologous relationships between the 11 groups. Surprisingly, 10
607 groups exhibited homology with each other, with the exception of PG6, which includes IncP-6
608 plasmids. Based on the highly sensitive HMM–HMM comparisons using HHblits, we identified
609 two supergroups (SG1 and SG2). The groups within each SG shared homology (>99%
610 probability, evalue ≤1.5e-22) (Figure 5A). In addition, PG9, which includes IncP-9 plasmids,
611 shared homology (90.1 – 98.4% probability, evalue ≤ 8.8e-05) with all other groups except
612 PG6. The solved and predicted crystal structures of SG1 RIPs, such as RepA from plasmid pPS10
613 (PG7), RepE from plasmid F (PG7), π from plasmid R6K (PG7), and TrfA protein from plasmid
614 RK2 (PG1), contain two winged-helix domains (WH1 and WH2) (104). Although the crystal
615 structures of PG9, PG10, and PG5 RIPs have not been reported, based on the alphafold3
616 predictions, these are predicted to have only one WH domain (Figure 2). We observed that the
617 C-terminal WH domain (WH2) of SG1 RIPs shared homology with the WH domain of PG9 RIPs.
618 WH2 of the RepA protein of pPS10 binds precisely to the iteron 3' end and is the major
619 determinant of RepA sequence specificity (105). We therefore speculated that the conserved
620 WH domain likely plays an important role in replication origin recognition. Here, we named
621 these WH domain-containing RIPs "WH RIPs", which constitute a "WH RIP superfamily".

622 The remaining PG6 exhibited no homology with the other groups (<1% probability by HHblits).
623 To characterise the evolutionary origin of PG6, we performed a HHblits search against Pfam
624 v35.0. The RIPs of PG6 show significant homology to PF03090 (Replicase family; with 98.4%
625 probability by HHblits), including the ColE2 Rep protein. The ColE2 Rep protein belongs to the
626 superfamily of AEPs (106) and is the only plasmid-specific trans-acting factor required for
627 initiation of plasmid replication (107). The RIPs of IncP-6 plasmids (9, 89) have a PriCT domain,

628 which may form a compact module and play a central role in DNA unwinding. The C termini of
629 some AEPs suggest that it probably plays a similar role in these proteins. The RIPs of PG6
630 plasmids have domain compositions different from those of the other groups. We therefore
631 concluded that RIPs of PG6 have different evolutionary origins with the other groups.

632 We reconstructed the phylogeny of the WH RIPs using the identified homologous region of the
633 WH domain. The homologous region showed similar secondary structure compositions
634 including multiple alpha helixes and beta turns among the 10 groups of the WH RIPs (Figure
635 5B). A maximum-likelihood phylogenetic tree of the WH RIPs was reconstructed, which
636 comprised 6,194 clustering representatives (60% identity) associated with 28,424 PLSDB
637 plasmids and 66,527 Uniparc proteins (Figure 5C). The tree is subdivided largely according to
638 the 10 groups with two special notes. First, PG7 (including IncP-7 plasmids) and PGsty
639 (including pSTY plasmids) were merged and formed one large group because these are being
640 composed of similar sequences. Second, PG5 (including IncP-5 plasmids) formed a subclade
641 within the clade of PG10. We defined clades considering the group affiliations and branch
642 support values. Most clades (i.e., PG5, PG9, PG16, PG3, PG15, PG4, and PG7/PGsty) were
643 identified with two types of high support values (UFBoot2: 72–100%, TBE: 99–100%). The
644 remaining (PG1 and PG10) clades were not highly supported by UFBoot2 but were supported
645 by TBE (>99%). Considering the larger structure, we defined a superclade PG1/PG4/PG15,
646 containing the PG1, PG4, and PG15 clades, and a superclade PG5/PG9/PG10, containing the
647 PG5, PG9, and PG10 clades, both of which were supported with high support values (UFBoot2:
648 77%, TBE: 99.9–100%). Notably, the branch dividing the PG5/PG9/PG10 superclade and the
649 remaining tree constitutes the well-supported division of SG1, which is composed of two WH
650 domains, and SG2/PG9, which is composed of one WH domain. We thus speculated that this
651 division is possibly the most ancestral split in the tree.

652 The tree includes diverse sequences not derived from PLSDB plasmids, indicating that
653 numerous uncharacterised plasmids and incompatibility groups exist. In total, 75.4% (n =
654 4,669) of tree leaves were not associated with PLSDB plasmids (Figure 5C). Conversely, a
655 substantial proportion of PLSDB plasmids (47.4%; n = 28,424) was associated with the
656 remaining leaves of the tree (n = 1,525), indicating that these plasmids are likely replicated by
657 WH RIPs. Considering plasmids with alternative replication mechanisms that employ disparate
658 types of RIPs, such as the superfamily of AEPs (e.g., IncP-6) and rolling-circle replication (e.g.,
659 Pfam PF01719; Rep_OBD), this fraction is notably high, almost reaching 29,662 PLSDB plasmids
660 identified with plasmid types using PlasmidFinder v2.1
661 (https://figshare.com/s/915a336484ee084ffbea; 59,895 plasmid list). Notably, 9,666 plasmids
662 lacking PlasmidFinder types were associated with the tree. In other words, besides the 2,357
663 PIncs plasmids, 8,860 additional plasmids were associated with the tree. This indicates that the
664 tree can serve as a reference for the characterisation of plasmids, facilitating the identification
665 and classification of RIP sequences. Furthermore, among the 1,525 leaves associated with
666 PLSDB plasmids, 51 leaves are notably associated with a large number of plasmids, each linked
667 to more than 100 plasmids, with a range from 102 to 6,872 plasmids, and collectively
668 accounting for 32,713 RIPs, which represents 81.0% of the total 40,371 RIPs. This suggests a

669 biased knowledge of plasmids containing similar RIPs. Subsequently, 9,141 plasmids (32.2%)
670 were identified with multiple RIPs, suggesting that multi-replicons are not a special case of
671 plasmids.

672 Each clade of the tree contains diverse taxonomic lineages, suggesting potential host
673 transitions within the clade (Figure 5C). In contrast, some taxa are enriched in specific clades.
674 For instance, the phylum Actinomycetota is enriched in PG15 (84.2%), whereas the phyla
675 Bacillota, Bacteroidota, Campylobacterota, and Fusobacteriota are enriched in PG7/PGsty
676 (90.0, 93.0, 95.0, 100%, respectively). Although this enrichment might be associated with some
677 specific environment (e.g., gut microbe for the latter), further study is required to describe the
678 detailed association between clades, lineages, and environments.

679 In summary, remote homology detection and phylogenetic reconstruction of the RIPs of PIncs
680 and their homologs lead to the following findings: (i) except IncP-6, the RIPs of PIncs exhibit
681 homology and are classified as members of the WH RIP superfamily, (ii) a well-supported
682 branch in the phylogenetic tree divided WH RIPs into the PG5/PG9/PG10 superclade, which
683 comprises one WH domain, and the remaining WH RIPs, which are composed of two WH
684 domains, and (iii) the tree greatly expanded the plasmid universe through the identification
685 and classification of the WH RIP superfamily.

686 **CONCLUSION**

687 This study provides valuable insights into the classification and genetic characterisation of
688 *Pseudomonas* plasmids. We successfully determined the complete whole nucleotide
689 sequences of several previously uncharacterised plasmids, including those in the
690 *Pseudomonas*, namely the IncP-5, IncP-11, and IncP-12 groups, and reclassified plasmids that
691 were previously thought to belong to 'IncP-13' plasmids as an ICE. Furthermore, we identified
692 RIP genes associated with these plasmids, along with the IncP-2 plasmid, contributing to a
693 more refined understanding of their replication mechanisms.

694 Our study also introduced four newly proposed plasmid groups, IncP-15, lncP-16, and
695 pQBR103-like and pSTY-like groups, thereby significantly expanding the plasmid groups and
696 contributing to the expansion and refinement of the *Pseudomonas* plasmid database. Among
697 the extensive collection of 59,895 plasmids in the PLSDB, a subset (2,351 plasmids) was
698 classified into these specific PIncs, which underlines the importance of updating and refining
699 plasmid databases to reflect new discoveries.

700 Our findings underscore the strengths of our classification system, which is grounded in
701 experimental data and aligns closely with established knowledge. In contrast, the recently
702 launched IMG/Plasmid database, despite its comprehensiveness in terms of including a vast
703 number of plasmid sequences (>700,000 sequences) (https://img.jgi.doe.gov/cgi-
704 bin/plasmid/main.cgi)(108), employs plasmid taxonomic units (PTUs) based solely on average
705 nucleotide sequence identity (109). Although this method is thorough in scope, it fails to

706 capture the distinct characteristics of individual plasmid groups that have been documented
707 through years of research as they are affected by long-sized accessory genes in plasmids. The
708 experimental identification of genes and DNA regions involved in plasmid replication remains
709 crucial for a deeper understanding of plasmid biology, including their host range, coevolution
710 with bacterial hosts, and potential applications in biotechnology. These findings are
711 particularly relevant for understanding the spread of ARG-associated plasmids in clinical and
712 environmental *Pseudomonas* isolates, especially in the context of multidrug-resistant *P.*
713 *aeruginosa* strains.

714 The reconstructed phylogenetic tree of WH RIPs will provide a fundamental basis for plasmid
715 detection and classification based on RIPs, which are essential proteins for understanding
716 plasmid replication mechanisms. Further research should focus on integrating experimental
717 data with large-scale sequence analyses to reconcile discrepancies between traditional and
718 modern classification systems. Additionally, experimental validation and phylogenetic analysis
719 of RIPs and *oriV* in plasmids in plasmids with alternative replication mechanisms, including the
720 superfamily of AEPs and rolling-circle replication, are essential. These efforts will enhance our
721 ability to track and manage antimicrobial resistance in various environments, ultimately
722 contributing to better public health outcomes.

723

724 **DATA AVAILABILITY**

725 The annotated plasmid sequences have been deposited in the GenBank/EMBL/DDBJ under
726 accession numbers LC685025, LC685026, LC685027, LC685593, and LC700336.

727

728 **Supplementary Data statement**
729 Supplementary Data are available at online.

730

731 **AUTHOR CONTRIBUTIONS**

732 The YN, HS, MS, MS conceived, designed, and supervised the study. KK, TK, MT, HX, YT, AH,
733 RM, HD, MS, HS, MS performed the experiments and data analysis. YN, RM, HD, KK, HN, MS,
734 HS, MS wrote, reviewed, and edited the manuscript. All authors contributed to manuscript
735 revision, read and approved the submitted version.

736

737 **ACKNOWLEDGEMENTS**

738 We are grateful to Prof. Dr. S. Iyobe of the Gumma University School of Medicine and Prof. Dr.
739 M. Tsuda of Tohoku University for providing Rms139, Rms163, RP1-1, R716, and pMG26.

Computation time was provided by the Super Computer System, Institute for Chemical Research, Kyoto University.

FUNDING

This work was supported by grants (JP24wm0225029 to Y. Tsuda; JP24gm1610007 to A. Hirabayashi; JP24fk0108665, JP24fk0108683, JP24fk0108712, JP24fk0108642, JP24gm1610003, JP24wm0225029, and JP24wm0225022 to M. Suzuki; JP22wm0325022 and JP22wm0225008 to M. Shintani) from the Japan Agency for Medical Research and Development (AMED), grants (JPMJAX21BK to Y. Nishimura; 19H02869 and 19H05679 to H. Nojiri; 20K10436 and JPMJCR20H1 to H. Suzuki; 22KK0058 and 23K06556 to M. Suzuki; 16H06279, 19H02869, 19H05686, and JP23H02124 to M. Shintani) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, grants to Y. Nishimura, M. Tokuda, X. Hui, C. Suzuki-Minakuchi, H. Nojiri, H. Suzuki, M. Suzuki, and M. Shintani from Consortium for the Exploration of Microbial Functions of Ohsumi Frontier Science Foundation, Japan, a grant from Asahi Glass Foundation to M. Shintani, a grant to M. Shintani from Institute of Fermentation Osaka (L-2023-1-002), National University Corporation Shizuoka University, Japan, and a grant to M. Shintani from the Research Institute of Green Science and Technology Fund for Research Project Support (2023-RIGST-23104), National University Corporation Shizuoka University, Japan.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Helinski,D.R. (2022) A Brief History of Plasmids. *EcoSal Plus*, **10**, eESP00282021.

2. Boolchandani,M., D'Souza,A.W. and Dantas,G. (2019) Sequencing-based methods and resources to study antimicrobial resistance. *Nat. Rev. Genet.*, **20**, 356–370.

3. Suzuki,M., Hashimoto,Y., Hirabayashi,A., Yahara,K., Yoshida,M., Fukano,H., Hoshino,Y., Shibayama,K. and Tomita,H. (2023) Genomic epidemiological analysis of antimicrobial-resistant bacteria with nanopore sequencing. In *Methods in Molecular Biology*, Methods in molecular biology (Clifton, N.J.). Springer US, New York, NY, pp. 227–246.

4. Novick,R.P. (1987) Plasmid incompatibility. *Microbiol. Rev.*, **51**, 381–395.

5. Miller,W.R. and Arias,C.A. (2024) ESKAPE pathogens: antimicrobial resistance, epidemiology, clinical impact and therapeutics. *Nat. Rev. Microbiol.*, 10.1038/s41579-024-01054-w.

774 6. Peix,A., Ramírez-Bahena,M.-H. and Velázquez,E. (2018) The current status on the taxonomy
775 of Pseudomonas revisited: An update. *Infect. Genet. Evol.*, **57**, 106–116.

776 7. Thomas,C.M. and Haines,A.S. (2004) Plasmids of the Genus Pseudomonas. In Ramos,J.-L.
777 (ed), *Pseudomonas: Volume 1 Genomics, Life Style and Molecular Architecture*.
778 Springer US, Boston, MA, pp. 197–231.

779 8. Lawley,T., Wilkins,B.M. and Frost,L.S. (2014) Bacterial conjugation in gram-negative bacteria.
780 In *Plasmid Biology*. ASM Press, Washington, DC, USA, pp. 203–226.

781 9. Haines,A.S., Cheung,M. and Thomas,C.M. (2006) Evidence that IncG (IncP-6) and IncU
782 plasmids form a single incompatibility group. *Plasmid*, **55**, 210–215.

783 10. Fernandez-Alarcon,C., Singer,R.S. and Johnson,T.J. (2011) Comparative genomics of
784 multidrug resistance-encoding IncA/C plasmids from commensal and pathogenic
785 Escherichia coli from multiple animal sources. *PLoS One*, **6**, e23415.

786 11. Yano,H., Shintani,M., Tomita,M., Suzuki,H. and Oshima,T. (2019) Reconsidering plasmid
787 maintenance factors for computational plasmid design. *Comput. Struct. Biotechnol. J.*,
788 **17**, 70–81.

789 12. Loftie-Eaton,W. and Rawlings,D.E. (2012) Diversity, biology and evolution of IncQ-family
790 plasmids. *Plasmid*, **67**, 15–34.

791 13. Xiong,J., Alexander,D.C., Ma,J.H., Deraspe,M., Low,D.E., Jamieson,F.B. and Roy,P.H. (2013)
792 Complete sequence of pOZ176, a 500-kilobase IncP-2 plasmid encoding IMP-9-
793 mediated carbapenem resistance, from outbreak isolate Pseudomonas aeruginosa 96.
794 *Antimicrob. Agents Chemother.*, **57**, 3775–3782.

795 14. Cazares,A., Moore,M.P., Hall,J.P.J., Wright,L.L., Grimes,M., Emond-Rhéault,J.-G.,
796 Pongchaikul,P., Santanirand,P., Levesque,R.C., Fothergill,J.L., *et al.* (2020) A
797 megaplasmid family driving dissemination of multidrug resistance in Pseudomonas.
798 *Nat. Commun.*, **11**, 1370.

799 15. Jiang,X., Yin,Z., Yuan,M., Cheng,Q., Hu,L., Xu,Y., Yang,W., Yang,H., Zhao,Y., Zhao,X., *et al.*
800 (2020) Plasmids of novel incompatibility group IncpRBL16 from Pseudomonas species.
801 *J. Antimicrob. Chemother.*, **75**, 2093–2100.

802 16. Urbanowicz,P., Bitar,I., Izdebski,R., Baraniak,A., Literacka,E., Hrabák,J. and Gniadkowski,M.
803 (2021) Epidemic Territorial Spread of IncP-2-Type VIM-2 Carbapenemase-Encoding
804 Megaplasmids in Nosocomial Pseudomonas aeruginosa Populations. *Antimicrob.*
805 *Agents Chemother.*, **65**.

806 17. Jacoby,G.A. (1986) Resistance Plasmids of Pseudomonas. *The Biology of Pseudomonas*,
807 10.1016/b978-0-12-307210-8.50013-0.

808 18. Sawada,Y., Yaginuma,S., Tai,M., Iyobe,S. and Mitsuhashi,S. (1976) Plasmid-mediated
809 penicillin beta-lactamases in Pseudomonas aeruginosa. *Antimicrob. Agents*
810 *Chemother.*, **9**, 55–60.

811 19. Sagai,H., Hasuda,K., Iyobe,S., Bryan,L.E., Holloway,B.W. and Mitsuhashi,S. (1976)
812 Classification of R plasmids by incompatibility in Pseudomonas aeruginosa. *Antimicrob.*
813 *Agents Chemother.*, **10**, 573–578.

814 20. Summers,A.O. and Jacoby,G.A. (1978) Plasmid-determined resistance to boron and
815 chromium compounds in Pseudomonas aeruginosa. *Antimicrob. Agents Chemother.*,
816 **13**, 637–640.

817 21. Ingram,L., Sykes,R.B., Grinsted,J., Saunders,J.R. and Richmond,M.H. (1972) A transmissible
818 resistance element from a strain of Pseudomonas aeruginosa containing no detectable
819 extrachromosomal DNA. *J. Gen. Microbiol.*, **72**, 269–279.

820 22. Bryan,L.E., Semaka,S.D., Van den Elzen,H.M., Kinnear,J.E. and Whitehouse,R.L. (1973)
821 Characteristics of R931 and other Pseudomonas aeruginosa R factors. *Antimicrob.*
822 *Agents Chemother.*, **3**, 625–637.

823 23. Jacoby,G.A. (1980) Plasmid determined resistance to carbenicillin and gentamicin in
824 *Pseudomonas aeruginosa*. *Plasmids and Transposons*, 10.1016/b978-0-12-675550-
825 3.50011-3.

826 24. Bradley,D.E. (1983) Specification of the conjugative pili and surface mating systems of
827 Pseudomonas plasmids. *J. Gen. Microbiol.*, **129**, 2545–2556.

828 25. Cain,D. and Holloway,B.W. (1984) Prime plasmids derived from the IncP-10 plasmid R91-5
829 in Pseudomonas putida. *FEMS Microbiol. Lett.*, **24**, 97–101.

830 26. Davies,S. and Krishnapillai,V. (1990) DNA sequence analysis of the replication region of the
831 *Pseudomonas aeruginosa* plasmid R91-5. *J. Genet.*, **69**, 101–112.

832 27. Schmartz,G.P., Hartung,A., Hirsch,P., Kern,F., Fehlmann,T., Müller,R. and Keller,A. (2022)
833 PLSDB: advancing a comprehensive database of bacterial plasmids. *Nucleic Acids Res.*,
834 **50**, D273–D278.

835 28. Galata,V., Fehlmann,T., Backes,C. and Keller,A. (2018) PLSDB: a resource of complete
836 bacterial plasmids. *Nucleic Acids Res.*, **47**, D195-D202.

837 29. Sambrook,J. and Russell,D. (2001) Molecular Cloning: A Laboratory Manual 3rd edn. Cold
838 Spring Harbor Laboratory, Cold Spring Harbor, NY.

839 30. Shintani,M., Yoshida,T., Habe,H., Omori,T. and Nojiri,H. (2005) Large plasmid pCAR2 and
840 class II transposon Tn*4676* are functional mobile genetic elements to distribute the
841 carbazole/dioxin-degradative *car* gene cluster in different bacteria. *Appl. Microbiol.*
842 *Biotechnol.*, **67**, 370–382.

843 31. Royle,P.L., Matsumoto,H. and Holloway,B.W. (1981) Genetic circularity of the
844 *Pseudomonas aeruginosa* PAO chromosome. *J. Bacteriol.*, **145**, 145–155.

845 32. Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA. (2016) plasmidSPAdes:
846 assembling plasmids from whole genome sequencing data. *Bioinformatics*, **32**, 3380-
847 3387.

848 33. Kearse,M., Moir,R., Wilson,A., Stones-Havas,S., Cheung,M., Sturrock,S., Buxton,S.,
849 Cooper,A., Markowitz,S., Duran,C., *et al.* (2012) Geneious Basic: an integrated and
850 extendable desktop software platform for the organization and analysis of sequence
851 data. *Bioinformatics*, **28**, 1647–1649.

852 34. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G.,
853 Durbin,R. and 1000 Genome Project Data Processing Subgroup (2009) The sequence
854 alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

855 35. Shen,W., Le,S., Li,Y. and Hu,F. (2016) SeqKit: a cross-platform and ultrafast toolkit for
856 FASTA/Q file manipulation. *PLoS One*, **11**, e0163962.

857 36. Bankevich,A., Nurk,S., Antipov,D., Gurevich,A.A., Dvorkin,M., Kulikov,A.S., Lesin,V.M.,
858 Nikolenko,S.I., Pham,S., Prjibelski,A.D., *et al.* (2012) SPAdes: a new genome assembly
859 algorithm and its applications to single-cell sequencing. *J. Comput. Biol.*, **19**, 455–477.

860 37. Tanizawa,Y., Fujisawa,T. and Nakamura,Y. (2018) DFAST: a flexible prokaryotic genome
861 annotation pipeline for faster genome publication. *Bioinformatics*, **34**, 1037–1039.

862 38. Sullivan,M.J., Petty,N.K. and Beatson,S.A. (2011) Easyfig: a genome comparison visualizer.
863 *Bioinformatics*, **27**, 1009–1010.

864 39. Florensa,A.F., Kaas,R.S., Clausen,P.T.L.C., Aytan-Aktug,D. and Aarestrup,F.M. (2022)
865 ResFinder - an open online resource for identification of antimicrobial resistance genes
866 in next-generation sequencing data and prediction of phenotypes from genotypes.
867 *Microb Genom*, **8**, 000748.

868 40. Bharat,A., Petkau,A., Avery,B.P., Chen,J.C., Folster,J.P., Carson,C.A., Kearney,A., Nadon,C.,
869 Mabon,P., Thiessen,J., *et al.* (2022) Correlation between phenotypic and *in silico*
870 detection of antimicrobial resistance in *Salmonella enterica* in Canada using Staramr.
871 *Microorganisms*, **10**, 292.

872 41. Feldgarden,M., Brover,V., Gonzalez-Escalona,N., Frye,J.G., Haendiges,J., Haft,D.H.,
873 Hoffmann,M., Pettengill,J.B., Prasad,A.B., Tillman,G.E., *et al.* (2021) AMRFinderPlus
874 and the reference gene catalog facilitate examination of the genomic links among
875 antimicrobial resistance, stress response, and virulence. *Sci. Rep.*, **11**, 12728.

876 42. Lobry,J.R. (1996) Origin of replication of *Mycoplasma genitalium*. *Science*, **272**, 745–746.

877 43. Necşulea,A. and Lobry,J.R. (2007) A new method for assessing the effect of replication on
878 DNA base composition asymmetry. *Mol. Biol. Evol.*, **24**, 2169–2179.

879 44. Solovyev,V. (2011) V. Solovyev, A Salamov (2011) Automatic annotation of microbial
880 genomes and metagenomic sequences. In metagenomics and its applications in
881 agriculture, biomedicine and environmental studies (Ed. R.W. Li), Nova Science
882 Publishers, p.61-78.

883 45. Abramson,J., Adler,J., Dunger,J., Evans,R., Green,T., Pritzel,A., Ronneberger,O., Willmore,L.,
884 Ballard,A.J., Bambrick,J., *et al.* (2024) Accurate structure prediction of biomolecular
885 interactions with AlphaFold 3. *Nature*, **630**, 493–500.

886 46. van Kempen,M., Kim,S.S., Tumescheit,C., Mirdita,M., Lee,J., Gilchrist,C.L.M., Söding,J. and
887 Steinegger,M. (2024) Fast and accurate protein structure search with Foldseek. *Nat.*
888 *Biotechnol.*, **42**, 243–246.

889 47. Swan,M.K., Bastia,D. and Davies,C. (2006) Crystal structure of π initiator protein–iteron
890 complex of plasmid R6K: Implications for initiation of plasmid DNA replication. *Proc.*
891 *Natl. Acad. Sci. U. S. A*, **103**, 18481–18486.

892 48. Schumacher,M.A., Tonthat,N.K., Kwong,S.M., Chinnam,N.B., Liu,M.A., Skurray,R.A. and
893 Firth,N. (2014) Mechanism of staphylococcal multiresistance plasmid replication origin
894 assembly by the RepA protein. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 9121–9126.

895 49. Robertson,J. and Nash,J.H.E. (2018) MOB-suite: software tools for clustering,
896 reconstruction and typing of plasmids from draft assemblies. *Microb. Genom.*, **4**,
897 e000206.

898 50. Carattoli,A. and Hasman,H. (2020) PlasmidFinder and in silico pMLST: Identification and
899 typing of plasmid replicons in whole-genome sequencing (WGS). *Methods Mol. Biol.*,
900 **2075**, 285–294.

901 51. Bayliss,S.C., Thorpe,H.A., Coyle,N.M., Sheppard,S.K. and Feil,E.J. (2019) PIRATE: A fast and
902 scalable pangenomics toolbox for clustering diverged orthologues in bacteria.
903 *Gigascience*, **8**, giz119.

904 52. Buchfink,B., Reuter,K. and Drost,H.-G. (2021) Sensitive protein alignments at tree-of-life
905 scale using DIAMOND. *Nat. Methods*, **18**, 366–368.

906 53. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J.
907 (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search
908 programs. *Nucleic Acids Res.*, **25**, 3389–3402.

909 54. Fassler,J. and Cooper,P. (2011) BLAST Glossary, National Center for Biotechnology
910 Information (US), https://www.ncbi.nlm.nih.gov/sites/books/NBK62051

911 55. Steinegger,M. and Söding,J. (2017) MMseqs2 enables sensitive protein sequence searching
912 for the analysis of massive data sets. *Nat. Biotechnol.*, **35**, 1026–1028.

913 56. Woodcroft,B.J., Boyd,J.A. and Tyson,G.W. (2016) OrfM: a fast open reading frame predictor
914 for metagenomic data. *Bioinformatics*, **32**, 2702–2703.

915 57. Eddy,S.R. (2011) Accelerated Profile HMM Searches. *PLoS Comput. Biol.*, **7**, e1002195.

916 58. Steinegger,M., Meier,M., Mirdita,M., Vöhringer,H., Haunsberger,S.J. and Söding,J. (2019)
917 HH-suite3 for fast remote homology detection and deep protein annotation. *BMC*
918 *Bioinformatics*, **20**, 473.

919 59. Wang,J., Youkharibache,P., Zhang,D., Lanczycki,C.J., Geer,R.C., Madej,T., Phan,L., Ward,M.,
920 Lu,S., Marchler,G.H., *et al.* (2020) iCn3D, a web-based 3D viewer for sharing 1D/2D/3D
921 representations of biomolecular structures. *Bioinformatics*, **36**, 131–135.

922 60. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7:
923 improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.

924 61. Minh,B.Q., Schmidt,H.A., Chernomor,O., Schrempf,D., Woodhams,M.D., von Haeseler,A.
925 and Lanfear,R. (2020) IQ-TREE 2: New models and efficient methods for phylogenetic
926 inference in the genomic era. *Mol. Biol. Evol.*, **37**, 1530–1534.

927 62. Kalyaanamoorthy,S., Minh,B.Q., Wong,T.K.F., von Haeseler,A. and Jermiin,L.S. (2017)
928 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods*,
929 **14**, 587–589.

930 63. Hoang,D.T., Chernomor,O., von Haeseler,A., Minh,B.Q. and Vinh,L.S. (2018) UFBoot2:
931 improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.*, **35**, 518–522.

932 64. Lemoine,F., Domelevo Entfellner,J.-B., Wilkinson,E., Correia,D., Dávila Felipe,M., De
933 Oliveira,T. and Gascuel,O. (2018) Renewing Felsenstein's phylogenetic bootstrap in the
934 era of big data. *Nature*, **556**, 452–456.

935 65. Mai,U. and Mirarab,S. (2018) TreeShrink: fast and accurate detection of outlier long
936 branches in collections of phylogenetic trees. *BMC Genomics*, **19**, 272.

937 66. Lemoine,F. and Gascuel,O. (2021) Gotree/Goalign: toolkit and Go API to facilitate the
938 development of phylogenetic workflows. *NAR Genom Bioinform*, **3**, lqab075.

939 67. Kawakami,Y., Mikoshiba,F., Nagasaki,S., Matsumoto,H. and Tazaki,T. (1972) Prevalence of
940 *Pseudomonas aeruginosa* strains possessing R factor in a hospital. *J. Antibiot.* , **25**,
941 607–609.

942 68. Hedges,R.W. and Jacob,A.E. (1975) Mobilization of plasmid-borne drug resistance
943 determinants for transfer from *Pseudomonas aeruginosa* to *Escherichia coli*. *Mol. Gen.*
944 *Genet.*, **140**, 69–79.

945 69. Kawalek,A., Kotecka,K., Modrzejewska,M., Gawor,J., Jagura-Burdzy,G. and Bartosik,A.A.
946 (2020) Genome sequence of *Pseudomonas aeruginosa* PAO1161, a PAO1 derivative
947 with the ICE*Pae1161* integrative and conjugative element. *BMC Genomics*, **21**, 14.

948 70. Klockgether,J., Reva,O., Larbig,K. and Tümmler,B. (2004) Sequence analysis of the mobile
949 genome island pKLC102 of *Pseudomonas aeruginosa* C. *J. Bacteriol.*, **186**, 518–534.

950 71. Johnson,C.M. and Grossman,A.D. (2015) Integrative and conjugative elements (ICEs): what
951 they do and how they work. *Annu. Rev. Genet.*, **49**, 577–601.

952 72. Bradley,D.E., Taylor,D.E. and Cohen,D.R. (1980) Specification of surface mating systems
953 among conjugative drug resistance plasmids in *Escherichia coli* K-12. *J. Bacteriol.*, **143**,
954 1466–1470.

955 73. Bradley,D.E. (1984) Characteristics and function of thick and thin conjugative pili
956 determined by transfer-derepressed plasmids of incompatibility groups I1, I2, I5, B, K
957 and Z. *J. Gen. Microbiol.*, **130**, 1489–1502.

958 74. Kitao,T., Kubori,T. and Nagai,H. (2022) Recent advances in structural studies of the
959 *Legionella pneumophila* Dot/Icm type IV secretion system. *Microbiol. Immunol.*, **66**,
960 67–74.

961 75. Yanagiya,K., Maejima,Y., Nakata,H., Tokuda,M., Moriuchi,R., Dohra,H., Inoue,K.,
962 Ohkuma,M., Kimbara,K. and Shintani,M. (2018) Novel self-transmissible and broad-
963 host-range plasmids exogenously captured from anaerobic granules or cow manure.
964 *Front. Microbiol.*, **9**, 2602.

965 76. Tokuda,M., Suzuki,H., Yanagiya,K., Yuki,M., Inoue,K., Ohkuma,M., Kimbara,K. and
966 Shintani,M. (2020) Determination of plasmid pSN1216-29 host range and the similarity
967 in oligonucleotide composition between plasmid and host chromosomes. *Front.*
968 *Microbiol.*, **11**, 1187.

969 77. Tokuda,M., Yuki,M., Ohkuma,M., Kimbara,K., Suzuki,H. and Shintani,M. (2023)
970 Transconjugant range of PromA plasmids in microbial communities is predicted by
971 sequence similarity with the bacterial host chromosome. *Microb. Genom.*, **9**,
972 mgen001043.

973 78. Hayakawa,M., Tokuda,M., Kaneko,K., Nakamichi,K., Yamamoto,Y., Kamijo,T., Umeki,H.,
974 Chiba,R., Yamada,R., Mori,M., *et al.* (2022) Hitherto-unnoticed self-transmissible
975 plasmids widely distributed among different environments in Japan. *Appl. Environ.*
976 *Microbiol.*, **88**, e0111422.

977 79. Lilley,A.K. and Bailey,M.J. (1997) Impact of plasmid pQBR103 acquisition and carriage on
978 the phytosphere fitness of *Pseudomonas fluorescens* SBW25: burden and benefit.
979 *Appl. Environ. Microbiol.*, **63**, 1584–1587.

980 80. Tett,A., Spiers,A.J., Crossman,L.C., Ager,D., Ciric,L., Dow,J.M., Fry,J.C., Harris,D., Lilley,A.,
981 Oliver,A., *et al.* (2007) Sequence-based analysis of pQBR103; a representative of a
982 unique, transfer-proficient mega plasmid resident in the microbial community of sugar
983 beet. *ISME J.*, **1**, 331–340.

984 81. Lv,L., Wan,M., Wang,C., Gao,X., Yang,Q., Partridge,S.R., Wang,Y., Zong,Z., Doi,Y., Shen,J., *et*
985 *al.* (2020) Emergence of a plasmid-encoded resistance-nodulation-division efflux pump
986 conferring resistance to multiple drugs, including tigecycline, in *Klebsiella pneumoniae*.
987 *mBio*, **11**, e02930-19.

988 82. Köhler,K.A.K., Rückert,C., Schatschneider,S., Vorhölter,F.-J., Szczepanowski,R., Blank,L.M.,
989 Niehaus,K., Goesmann,A., Pühler,A., Kalinowski,J., *et al.* (2013) Complete genome
990 sequence of *Pseudomonas* sp. strain VLB120 a solvent tolerant, styrene degrading
991 bacterium, isolated from forest soil. *J. Biotechnol.*, **168**, 729–730.

992 83. Chen,F., Wang,P., Yin,Z., Yang,H., Hu,L., Yu,T., Jing,Y., Guan,J., Wu,J. and Zhou,D. (2022)
993 VIM-encoding IncpSTY plasmids and chromosome-borne integrative and mobilizable
994 elements (IMEs) and integrative and conjugative elements (ICEs) in *Pseudomonas*. *Ann.*
995 *Clin. Microbiol. Antimicrob.*, **21**, 10.

996 84. Pachulec,E. and van der Does,C. (2010) Conjugative plasmids of *Neisseria gonorrhoeae*.
997 *PLoS One*, **5**, e9962.

85. Shintani,M., Suzuki,H., Nojiri,H. and Suzuki,M. (2022) Precise classification of antimicrobial resistance-associated IncP-2 megaplasmids for molecular epidemiological studies on *Pseudomonas* species. *J. Antimicrob. Chemother.*, 10.1093/jac/dkac006.

86. Hidaka,M., Kobayashi,T., Takenaka,S., Takeya,H. and Horiuchi,T. (1989) Purification of a DNA replication terminus (*ter*) site-binding protein in *Escherichia coli* and identification of the structural gene. *J. Biol. Chem.*, **264**, 21031–21037.

87. Ambrose,S.J., Harmer,C.J. and Hall,R.M. (2018) Compatibility and entry exclusion of IncA and IncC plasmids revisited: IncA and IncC plasmids are compatible. *Plasmid*, **96–97**, 7–12.

88. Robinson,L.A., Collins,A.C.Z., Murphy,R.A., Davies,J.C. and Allsopp,L.P. (2023) Diversity and prevalence of type VI secretion system effectors in clinical *Pseudomonas aeruginosa* isolates. *Front. Microbiol.*, **13**.

89. Haines,A.S., Jones,K., Cheung,M. and Thomas,C.M. (2005) The IncP-6 plasmid Rms149 consists of a small mobilizable backbone with multiple large insertions. *J. Bacteriol.*, **187**, 4728–4738.

90. Kulinska,A., Czeredys,M., Hayes,F. and Jagura-Burdzy,G. (2008) Genomic and functional characterization of the modular broad-host-range RA3 plasmid, the archetype of the IncU group. *Appl. Environ. Microbiol.*, **74**, 4119–4132.

91. Rhodes,G., Parkhill,J., Bird,C., Ambrose,K., Jones,M.C., Huys,G., Swings,J. and Pickup,R.W. (2004) Complete nucleotide sequence of the conjugative tetracycline resistance plasmid pFBAOT6, a member of a group of IncU plasmids with global ubiquity. *Appl. Environ. Microbiol.*, **70**, 7497–7510.

92. Shintani,M., Yano,H., Habe,H., Omori,T., Yamane,H., Tsuda,M. and Nojiri,H. (2006) Characterization of the replication, maintenance, and transfer features of the IncP-7 plasmid pCAR1, which carries genes involved in carbazole and dioxin degradation. *Appl. Environ. Microbiol.*, **72**, 3206–3216.

93. Yano,H., Garruto,C.E., Sota,M., Ohtsubo,Y., Nagata,Y., Zylstra,G.J., Williams,P.A. and Tsuda,M. (2007) Complete sequence determination combined with analysis of transposition/site-specific recombination events to explain genetic organization of IncP-7 TOL plasmid pWW53 and related mobile genetic elements. *J. Mol. Biol.*, **369**, 11–26.

94. Yano,H., Miyakoshi,M., Ohshima,K., Tabata,M., Nagata,Y., Hattori,M. and Tsuda,M. (2010) Complete nucleotide sequence of TOL plasmid pDK1 provides evidence for evolutionary history of IncP-7 catabolic plasmids. *J. Bacteriol.*, **192**, 4337–4347.

95. Heinaru,E., Truu,J., Stottmeister,U. and Heinaru,A. (2000) Three types of phenol and p-cresol catabolism in phenol- and p-cresol-degrading bacteria isolated from river water continuously polluted with phenolic compounds. *FEMS Microbiol. Ecol.*, **31**, 195–205.

1035 96. Lopez-Echartea,E., Suman,J., Smrhova,T., Ridl,J., Pajer,P., Strejcek,M. and Uhlik,O. (2021)
1036 Genomic analysis of dibenzofuran-degrading *Pseudomonas veronii* strain Pvy reveals
1037 its biodegradative versatility. *G3* , **11**.

1038 97. Suenaga,H., Fujihara,H., Kimura,N., Hirose,J., Watanabe,T., Futagami,T., Goto,M.,
1039 Shimodaira,J. and Furukawa,K. (2017) Insights into the genomic plasticity of
1040 *Pseudomonas putida* KF715, a strain with unique biphenyl-utilizing activity and
1041 genome instability properties. *Environ. Microbiol. Rep.*, **9**, 589–598.

1042 98. Sevastsyanovich,Y.R., Krasowiak,R., Bingle,L.E., Haines,A.S., Sokolov,S.L., Kosheleva,I.A.,
1043 Leuchuk,A.A., Titok,M.A., Smalla,K. and Thomas,C.M. (2008) Diversity of IncP-9
1044 plasmids of *Pseudomonas*. *Microbiology*, **154**, 2929–2941.

1045 99. Huang,W., Hamouche,J.E., Wang,G., Smith,M., Yin,C., Dhand,A., Dimitrova,N. and
1046 Fallon,J.T. (2020) Integrated genome-wide analysis of an isogenic pair of *Pseudomonas*
1047 *aeruginosa* clinical isolates with differential antimicrobial resistance to
1048 ceftolozane/tazoiactam, ceftazidime/avibactam, and piperacillin/tazobactam. *Int. J.*
1049 *Mol. Sci.*, **21**, 1026.

1050 100. Bonnin,R.A., Bogaerts,P., Girlich,D., Huang,T.-D., Dortet,L., Glupczynski,Y. and Naas,T.
1051 (2018) Molecular characterization of OXA-198 carbapenemase-producing
1052 *Pseudomonas aeruginosa* clinical isolates. *Antimicrob. Agents Chemother.*, **62**, e02496-
1053 17.

1054 101. Shintani,M., Suzuki,H., Nojiri,H. and Suzuki,M. (2023) Reconsideration of the previously
1055 classified incompatibility groups of plasmids, IncP-1 and IncP-11. *Environ. Microbiol.*,
1056 **25**, 1071-1076

1057 102. Botelho,J., Lood,C., Partridge,S.R., van Noort,V., Lavigne,R., Grosso,F. and Peixe,L. (2019)
1058 Combining sequencing approaches to fully resolve a carbapenemase-encoding
1059 megaplasmid in a *Pseudomonas shirazica* clinical strain. *Emerg. Microbes Infect.*, **8**,
1060 1186–1194.

1061 103. Zhu,J., Lv,J., Zhu,Z., Wang,T., Xie,X., Zhang,H., Chen,L. and Du,H. (2023) Identification of
1062 TMexCD-TOprJ-producing carbapenem-resistant Gram-negative bacteria from hospital
1063 sewage. *Drug Resist. Updat.*, **70**, 100989.

1064 104. Wegrzyn,K., Zabrocka,E., Bury,K., Tomiczek,B., Wieczor,M., Czub,J., Uciechowska,U.,
1065 Moreno-Del Alamo,M., Walkow,U., Grochowina,I., *et al.* (2021) Defining a novel
1066 domain that provides an essential contribution to site-specific interaction of Rep
1067 protein with DNA. *Nucleic Acids Res.*, **49**, 3394-3408.

1068 105. Giraldo,R. and Fernandez-Tresguerres,M.E. (2004) Twenty years of the pPS10 replicon:
1069 insights on the molecular mechanism for the activation of DNA replication in iteron-
1070 containing bacterial plasmids. *Plasmid*, **52**, 69–83.

1071 106. Iyer,L.M., Koonin,E.V., Leipe,D.D. and Aravind,L. (2005) Origin and evolution of the
1072 archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural
1073 insights and new members. *Nucleic Acids Res.*, **33**, 3875–3896.

1074 107. Itou,H., Yagura,M., Shirakihara,Y. and Itoh,T. (2015) Structural basis for replication origin
1075 unwinding by an initiator primase of plasmid ColE2-P9: duplex DNA unwinding by a
1076 single protein. *J. Biol. Chem.*, **290**, 3601–3611.

1077 108. Camargo,A.P., Roux,S., Schulz,F., Babinski,M., Xu,Y., Hu,B., Chain,P.S.G., Nayfach,S. and
1078 Kyrpides,N.C. (2023) Identification of mobile genetic elements with geNomad. *Nat.*
1079 *Biotechnol.*, **42**, 1303-1312

1080 109. Redondo-Salvo,S., Bartomeus-Peñalver,R., Vielva,L., Tagg,K.A., Webb,H.E., Fernández-
1081 López,R. and de la Cruz,F. (2021) COPLA, a taxonomic classifier of plasmids. *BMC*
1082 *Bioinformatics*, **22**, 390.

1083 110. Robertson,J. and Nash,J.H.E. (2018) MOB-suite: software tools for clustering,
1084 reconstruction and typing of plasmids from draft assemblies. *Microb. Genom.,* **4**,
1085 e000206.

1086 111. Holloway,B.W. (1955) Genetic recombination in *Pseudomonas aeruginosa*. *J. Gen.*
1087 *Microbiol.*, **13**, 572–581.

1088 112. Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P., Hickey,M.J.,
1089 Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M., *et al.* (2000) Complete genome
1090 sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*, **406**,
1091 959–964.

1092 113. Bagdasarian,M., Lurz,R., Rückert,B., Franklin,F.C., Bagdasarian,M.M., Frey,J. and
1093 Timmis,K.N. (1981) Specific-purpose plasmid cloning vectors. II. Broad host range, high
1094 copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in
1095 *Pseudomonas*. *Gene*, **16**, 237–247.

1096 114. Nelson,K.E., Weinel,C., Paulsen,I.T., Dodson,R.J., Hilbert,H., Martins dos Santos,V.A.,
1097 Fouts,D.E., Gill,S.R., Pop,M., Holmes,M., *et al.* (2002) Complete genome sequence and
1098 comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440.
1099 *Environ. Microbiol.*, **4**, 799–808.

1100 115. Holloway,B.W. and Richmond,M.H. (1973) R-factors used for genetic studies in strains of
1101 *Pseudomonas aeruginosa* and their origin. *Genet. Res.* , **21**, 103–105.

1102 116. Grinsted,J., Saunders,J.R., Ingram,L.C., Sykes,R.B. and Richmond,M.H. (1972) Properties of
1103 a R factor which originated in *Pseudomonas aeruginosa* 1822. *J. Bacteriol.*, **110**, 529–
1104 537.

1105

1106 **TABLES**

1107 **Table 1. Lists of representative plasmids in the PIncs.**

Incompatibility (Inc) groups in *Pseudomonas* (PIncs)	**Representative plasmid [accession no.]**	**MOB MPF[a]**	**No. of plasmids in PLSDB**	**Average size [bp]**	**Average GC [%]**	**No. of core genes[b]**	**No of multi-replicons [ratio(%)]**	**No. of plasmids with ARG [ratio(%)]**
IncP-1 (IncP)	RK2/RP4 [BN000925]	MOB_P MPF_T	321	57,024	55.9	25	13 [4.0]	166 [50.8]
IncP-2	**Rms139 [LC653116]**	MOB_P MPF_I	87	439,701	56.8	293	32 [36.8]	82 [96.7]
IncP-3A/C (IncA/C)	pRA1 (IncA) [FJ705807] pR55 (IncC) [JQ010984]	MOB_H MPF_F	921	165,307	52.0	48	211 [22.9]	891 [96.7]
IncP-4 (IncQ)	RSF1010 [M28829]	MOB_Q -	695	120,958	53.2	2	493 [70.9]	657 [94.5]
IncP-5	**Rms163**	NA	2	244,810	57.7	ND	0	1

	[LC685027]	NA					[0.0]	[50.0]
IncP-6	Rms149	MOB_P	119	42,711	58.7	3	72	105
(IncG)	[AJ877225]	-					[60.5]	[88.2]
IncP-7	pCAR1	MOB_H	23	156,431	56.0	10	2	4
	[AB088420]	MPF_F					[8.7]	[17.4]
IncP-9	pWW0 [AJ344068]	MOB_F	34	87,921	58.0	10	2	20
	NAH7 [AB237655]	MPF_F					[5.9]	[58.8]
	Rsu2 [LC685593]							
IncP-10	R91-5 [X54695[c]]	NA	49	49,589	59.0	19	25	30
	pPAB546	NA					[51.0]	[61.2]
	[MN433456]							
IncP-11	**RP1-1 [LC700336]**	NA	23	55,244	58.9	37	15	6
		NA					[65.2]	[26.1]

IncP-12	**R716** [**LC685026**]	NA NA	9	176,615	59.5	165	0 [0.0]	1 [11.1]
IncP-13	**pMG26d** **[LC685025]**	MOB_F MPF_G	NA	NA	NA	NA	NA	NA
IncP-15 (PromA)	pSN1104-11 [AP018707]	MOB_P MPF_T	42	50,043	61.1	30	1 [2.4]	3 [13.0]
IncP-16 (pSN1216-29)	pSN1216-29 [AP018710]	MOB_P MPF_T	24	51,023	60.1	17	4 [16.7]	6 [25.0]
unassigned (pQBR103)	pQBR103 [AM235768]	NA NA	11	514,821	55.1	229	1 [9.1]	7 [63.6]
Inc_{pSTY} (pSTY)	pSTY [NC_022739]	MOB_F MPF_F	15	297,214	56.5	66	4 [26.7]	8 [33.3]

1108 [a]Classification of MOB classes and MPF types was identified by the MOB-suite (110) with criteria of >=90% identity and >=60% coverage. "-"
1109 indicates that the genes involved in conjugation have not been detected, whereas "NA" indicates that the classes and types of MOB or MPF were
1110 not identified although the plasmid was self-transmissible. [b]Core genes of each Inc group plasmid were identified by PIRATE analyses. [c]Partial DNA
1111 sequences are available. [d]pMG26 was shown to be not a plasmid but an integrative and conjugative element in this study.

1112

1113 **Table 2. Bacterial strains and plasmids used in this study.**

Bacterial strains/ plasmids	Relevant characteristics	Reference
Bacterial strains		
Escherichia coli		
DH5α	F^-, *endA1*, *hsdR17*(r_k^-, m_k^-), *phoA*, *supE44*, *thi-1*, λ^-, *recA1*, *gyrA96*, *relA1*, Δ(*lacZYA- argF*)U169, ψ80d*lacZ*ΔM15	RBCBioscience
JM109	F' [*tra*D36, *proAB*, *lacI*^{q}, *lacZ*ΔM15], *recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17*(r_K^- m_K^+), e14$^-$ (*mcrA*$^-$), *supE44*, *relA1*, Δ(*lac-pro*AB)	RBCBioscience
Pseudomonas aeruginosa		
PAO1	Wild type isolate	(111, 112)
PAO1R	A spontaneous Rif^R-mutant	This study
PAO1808	A derivative strain of PAO1, *leu* auxotroph	(31)
P. putida KT2440	pWW0-free *Pseudomonas putida* mt-2	(113, 114)
Plasmids		
Rms139	IncP-2 plasmid, Cb^R, Cm^R, Sm^R, Tc^R	(18, 85)
Rms163	IncP-5 plasmid, Cm^R, Tc^R	(19, 20)

Rsu2	IncP-9 plasmid, Cb^R, Sm^R	(67, 68)
RP1-1	IncP-11 plasmid, Cb^R	(115, 116)
R716	IncP-12 plasmid, Sm^R	(22)
pMG26	Considered to be an IncP-13 plasmid, Cb^R, Cm^R, Km^R, Gm^R, Sm^R	(17)
pUC-GW-AMP	Cloning vector, Ap^R (Pip^R)	Azanta Life Science
mini-Rms139A (IncP-2)	pUC-GW-AMP carrying putative *oriV* and *repP-2A* regions of Rms139 (accession no. LC653116, 370,867-370,989 nt, 1-1181 nt and 2395-3150 nt)	This study
mini-Rms163A (IncP-5)	pUC-GW-AMP carrying putative *oriV* and *repP-5A* regions of Rms163 (accession no. LC685027, 267,484-267,599 nt and 1-1,506 nt)	This study
mini-pPAB546A (IncP-10)	pUC-GW-AMP carrying putative *oriV* and *repP-10A* regions of PAB546 plasmid (accession no. MN433456.1, 10,799-8,355 nt)	This study
mini-RP1-1A (IncP-11)	pUC-GW-AMP carrying putative *oriV* and *repP-11A* regions of RP1-1 (accession no. LC700336, 55,020-55,149 nt and 1-1,759 nt)	This study
mini-R716A (IncP-12)	pUC-GW-AMP carrying putative *oriV* and *repP-12A* regions of R716 (accession no. LC685026, 178,926-179,007 nt and 1-1,572 nt)	This study
mini-pMG26A (IncP-13)	pUC-GW-AMP carrying putative *oriV* and *rep* regions of pMG26 (accession no. LC685025, 1-1,344 nt and 10,751-12,119 nt)	This study

1114

Table 3. Conjugation frequency of each plasmid assessed with *Pseudomonas aeruginosa* strains PAO1 and KT2440 as recipients, using both filter (solid) mating and liquid mating assays.

Plasmid	Recipient			
	PAO1		KT2440	
	solid	liquid	solid	liquid
ICEpMG26 (Previous IncP-13)	$6.7 \pm 1.8 \times 10^{-2}$	$7.8 \pm 2.1 \times 10^{-3}$	$1.6 \pm 0.4 \times 10^{-4}$	$2.9 \pm 1.5 \times 10^{-6}$
Rms139 (IncP-2)	$1.1 \pm 0.78 \times 10^{-5}$	$8.0 \pm 2.9 \times 10^{-7}$	$1.6 \pm 0.44 \times 10^{-5}$	$4.7 \pm 1.5 \times 10^{-8}$
Rms163 (IncP-5)	$1.2 \pm 0.55 \times 10^{-4}$	$<2.0 \times 10^{-9}$	Not determined	Not determined
Rsu2 (IncP-9)	$2.8 \pm 1.4 \times 10^{-6}$	$8.8 \pm 2.2 \times 10^{-7}$	$8.6 \pm 0.79 \times 10^{-1}$	$1.5 \pm 0.34 \times 10^{0}$
RP1-1 (IncP-11)	$2.6 \pm 1.4 \times 10^{-4}$	$1.4 \pm 0.69 \times 10^{-5}$	Not determined	Not determined
R716 (IncP-12)	$9.5 \pm 2.6 \times 10^{-7}$	$4.8 \pm 1.3 \times 10^{-6}$	Not determined	Not determined

Figure 1. Genetic structure of mini-replicon of IncP-2 (Rms139), IncP-5 (Rms163), IncP-10 (pAP546), IncP-11 (RP1-1), and IncP-12 (R716) plasmids.
(A) The mini-replicon of each plasmid was constructed by using DNA regions shown in a red arrow (replication initiation protein RIP gene) with its promoter (grey box) and a grey solid line (*oriV*). Putative iterons, palindromes, and A+T-rich regions are shown by short red arrows, short yellow arrows, and blue boxes below the *oriV* region, respectively. Putative DnaA boxes are shown by black arrowheads in the *oriV* region. (B) Consensus sequences of putative iterons are shown. (C) Transformation efficiency of *P. aeruginosa* PAO1 with the mini-replicons are shown.

Figure 2. Predicted three-dimensional structure of RIP of plasmids in PIncs.
(A) Predicted structures of RIPs of IncP-1 (RepP-1A (=TrfA1) of RK2), IncP-2 (RepP-2A of Rms139), IncP-3 (RepP-3A of pRA1), IncP-4 (RepP-4A (=RepC) of RSF1010), IncP-7 (RepP-7A of pCAR1), IncP-9 (RepP-9A of pWW0), IncP-12 (RepP-12A of R716), IncP-15 (RepP-15A of pSN1104-11), IncP-16 (RepP-16A of pSB1216-29), pQBR103 ($Rep_{pQBR103}$), and pSTY (Rep_{pSTY}). The crystal structure of Pi protein of IncX plasmid, R6K, is shown as reference, of which structure is elucidated by X-ray crystallography. (B) Predicted structures of RIPs of IncP-5 (RepP-5A of Rms163), IncP-10 (RepP-10A of pPAB546), and

IncP-11 (RepP-11A of RP1-1). The crystal structure of Rep19A protein of *Staphylococcus* plasmid, pCA-347, is shown as reference, of which structure is elucidated by X-ray crystallography.

Figure 3. Features of plasmids of PIncs in the PLSDB.

Distribution of (A) plasmid sizes (kb) and (B) G+C contents (%), and (C) ratio of core and accessory genes of each PIncs plasmids in the PLSDB. (D) Relationship of each PIncs plasmids to other plasmid replicons and ARGs in the PLSDB. A sankey diagram was generated using the SankeyWidget library in Python through in-house programming.

Figure 4. Homologous relationships between 251 non-redundant RIPs of 15 PIncs.

%-identities of all-against-all BLASTp among the 251 non-redundant RIPs were visualised. The heatmap and the dendrograms were generated using R v3.6.1 and R packages dendextend and gplots via CPAP v0.1.0 (https://github.com/yosuken/CPAP) with a option ("--clust-method complete") to perform clustering with complete-linkage method.

Figure 5. Evolutionary relationships and phylogeny of WH RIPs.

(A) Remote homology among 10 WH RIP groups, identified using HHblits. Two supergroups (SG1 and SG2) were defined according to the high probability (>99%) predicted within each supergroup. Relationships with more than 90% probability were indicated by lines. The WH2 domain of SG1 is homologous to the WH domain of SG2 and PG9. The domain lengths shown are arbitrary. (B) The secondary structures of representatives of the 10 WH RIP groups. The representatives were selected from the 3D structure-resolved sequences shown in Figure 2. Namely, PG1 (RepP-1A (=TrfA1) of RK2), PG3 (RepP-3A of pRA1), PG4 (RepP-4A (=RepC) of RSF1010), PG7 (RepP-7A of pCAR1), PG15 (RepP-15A of pSN1104-11), PG16 (RepP-16A of pSB1216-29), PGsty (ReppSTY), PG9 (RepP-9A of pWW0), PG5 (RepP-5A of Rms163), and PG10 (RepP-10A of pPAB546) were selected. The secondary structures were predicted and visualised using iCn3D based on the 3D structures. (C) The ML phylogenetic tree of WH RIPs (6,194 leaves) calculated based on the conserved region (98 positions in the alignment) of the WH domain. The tree was rooted by the superclade of PG5/PG9/PG10. The clade and superclade names, bootstrap values, and number of leaves are shown in the legend. The consensus taxonomic affiliation of each leaf is shown in the outer ring. The number of leaves for each taxon is shown in the legend. The number of sequences (PLSDB plasmids and UniParc proteins) associated with each leaf is indicated by lines in the outmost rings. The line length represents the number of sequences but the upper limit in the visualisation is set to 100.

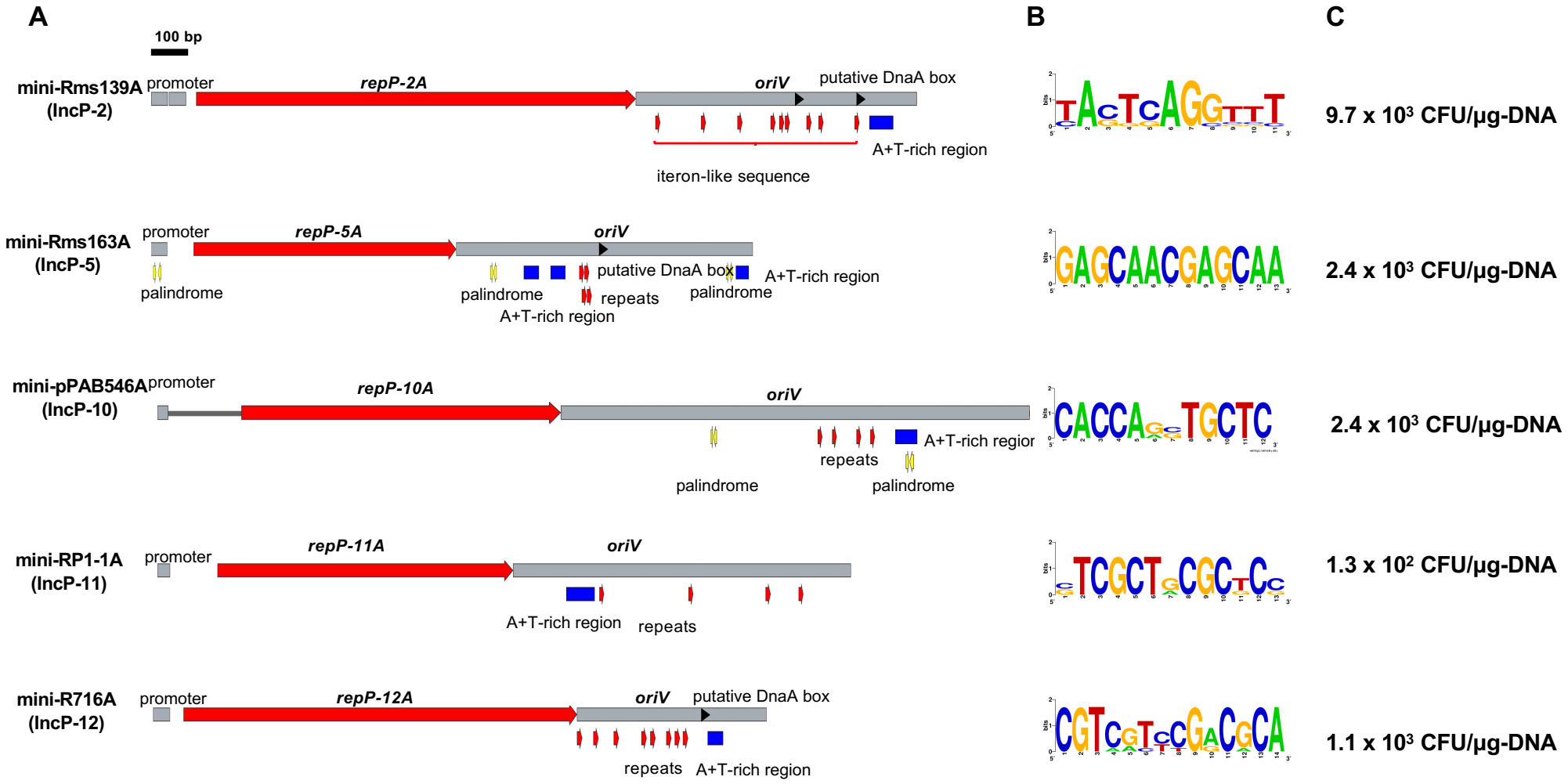
A
B
C
100 bp
mini-Rms139A (IncP-2)
promoter
repP-2A
oriV
putative DnaA box
A+T-rich region
iteron-like sequence
9.7 x 10³ CFU/µg-DNA
mini-Rms163A (IncP-5)
promoter
repP-5A
oriV
palindrome
palindrome
putative DnaA box
A+T-rich region
repeats
palindrome
A+T-rich region
2.4 x 10³ CFU/µg-DNA
mini-pPAB546A (IncP-10)
promoter
repP-10A
oriV
A+T-rich region
repeats
palindrome
palindrome
2.4 x 10³ CFU/µg-DNA
mini-RP1-1A (IncP-11)
promoter
repP-11A
oriV
A+T-rich region
repeats
1.3 x 10² CFU/µg-DNA
mini-R716A (IncP-12)
promoter
repP-12A
oriV
putative DnaA box
repeats
A+T-rich region
1.1 x 10³ CFU/µg-DNA

Fig. 1

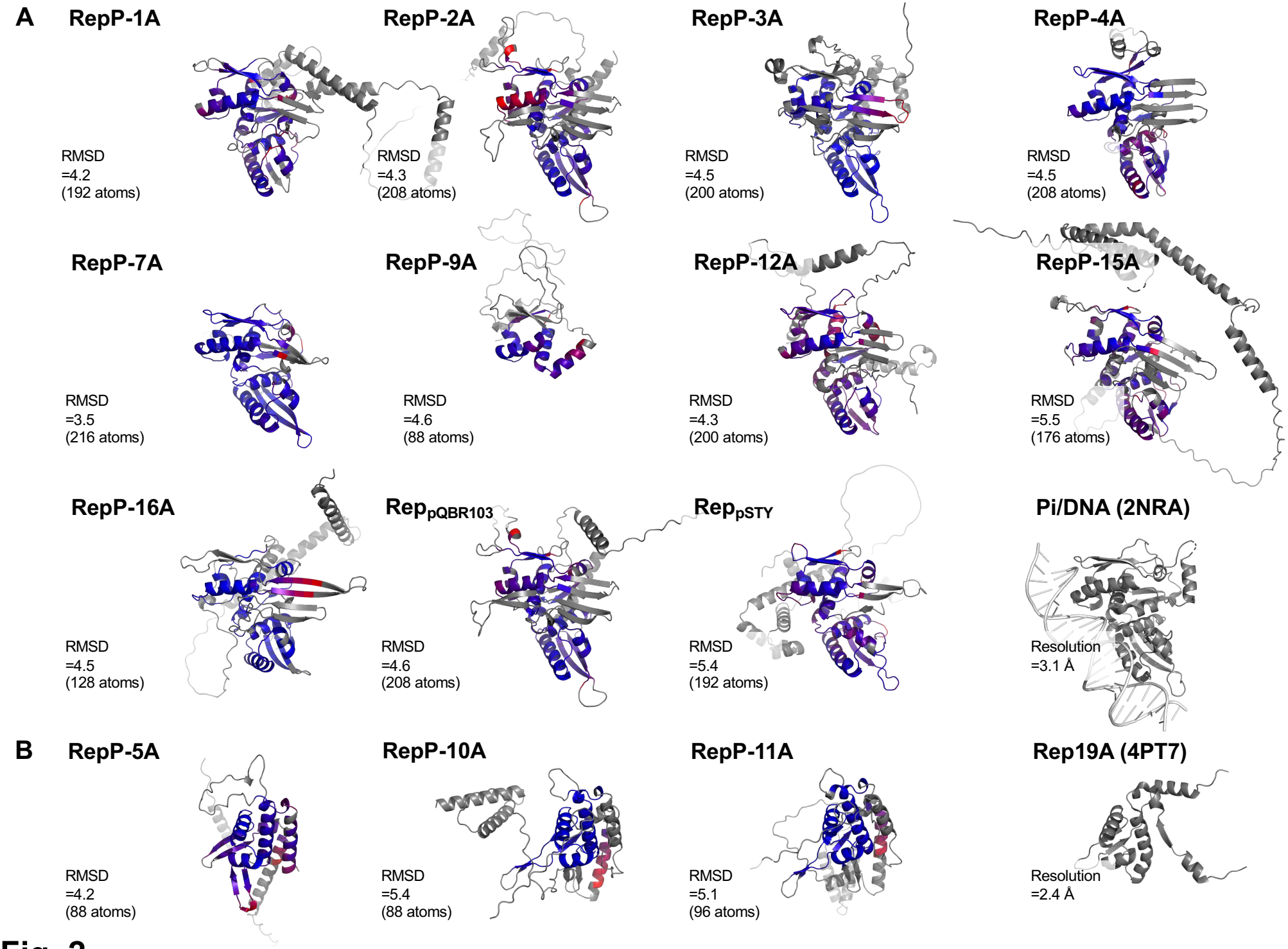
A
RepP-1A
RMSD =4.2 (192 atoms)
RepP-2A
RMSD =4.3 (208 atoms)
RepP-3A
RMSD =4.5 (200 atoms)
RepP-4A
RMSD =4.5 (208 atoms)
RepP-7A
RMSD =3.5 (216 atoms)
RepP-9A
RMSD =4.6 (88 atoms)
RepP-12A
RMSD =4.3 (200 atoms)
RepP-15A
RMSD =5.5 (176 atoms)
RepP-16A
RMSD =4.5 (128 atoms)
Rep pQBR103
RMSD =4.6 (208 atoms)
Rep pSTY
RMSD =5.4 (192 atoms)
Pi/DNA (2NRA)
Resolution =3.1 Å
B
RepP-5A
RMSD =4.2 (88 atoms)
RepP-10A
RMSD =5.4 (88 atoms)
RepP-11A
RMSD =5.1 (96 atoms)
Rep19A (4PT7)
Resolution =2.4 Å

Fig. 2

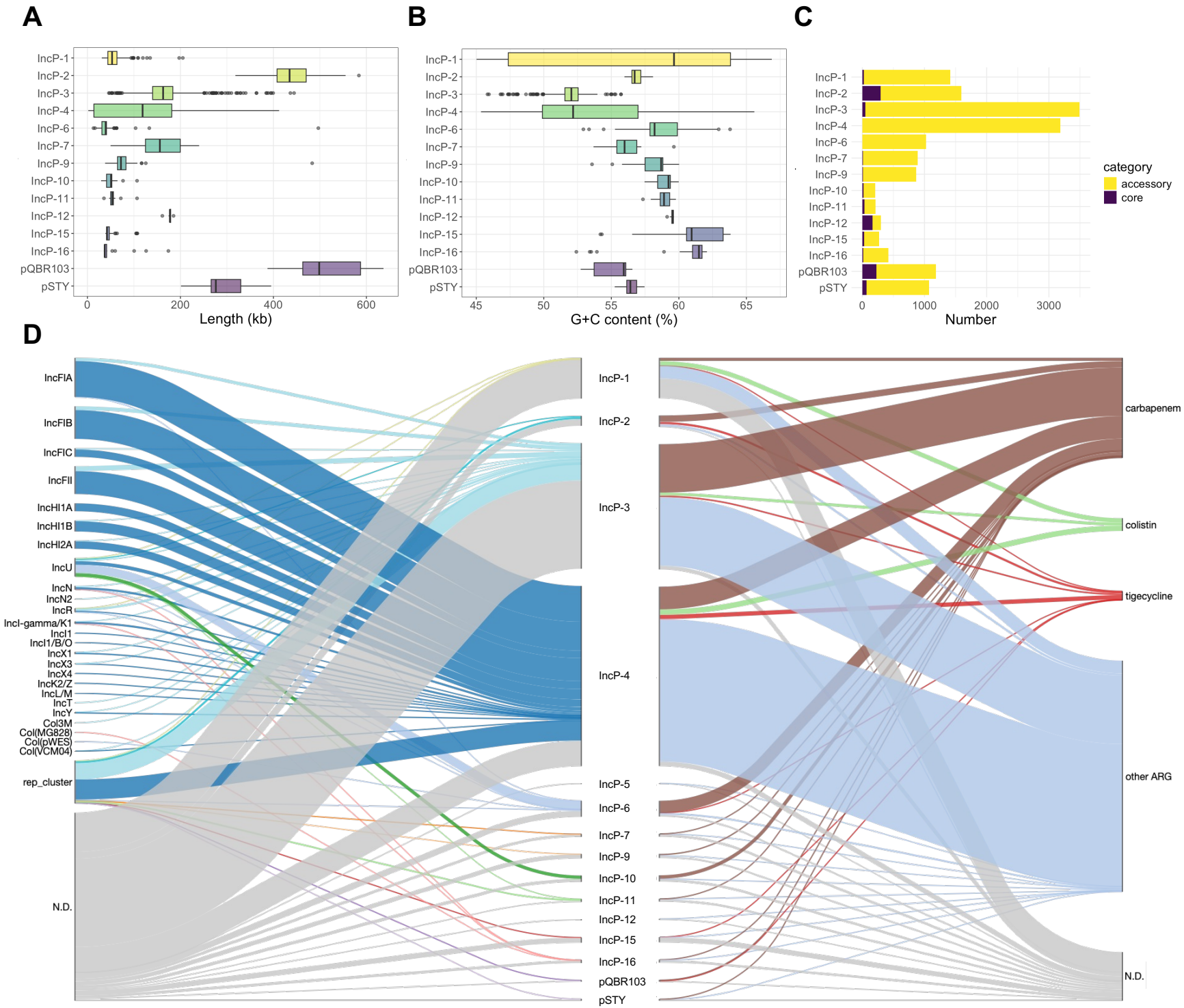
A
IncP-1
IncP-2
IncP-3
IncP-4
IncP-6
IncP-7
IncP-9
IncP-10
IncP-11
IncP-12
IncP-15
IncP-16
pQBR103
pSTY
0
200
400
600
Length (kb)
B
45
50
55
60
65
G+C content (%)
C
0
1000
2000
3000
Number
category
accessory
core
D
IncFIA
IncFIB
IncFIC
IncFII
IncHI1A
IncHI1B
IncHI2A
IncU
IncN
IncN2
IncR
IncI-gamma/K1
IncI1
IncI1/B/O
IncX1
IncX3
IncX4
IncK2/Z
IncL/M
IncT
IncY
Col3M
Col(MG828)
Col(pWES)
Col(VCM04)
rep_cluster
N.D.
carbapenem
colistin
tigecycline
other ARG
N.D.

Fig. 3

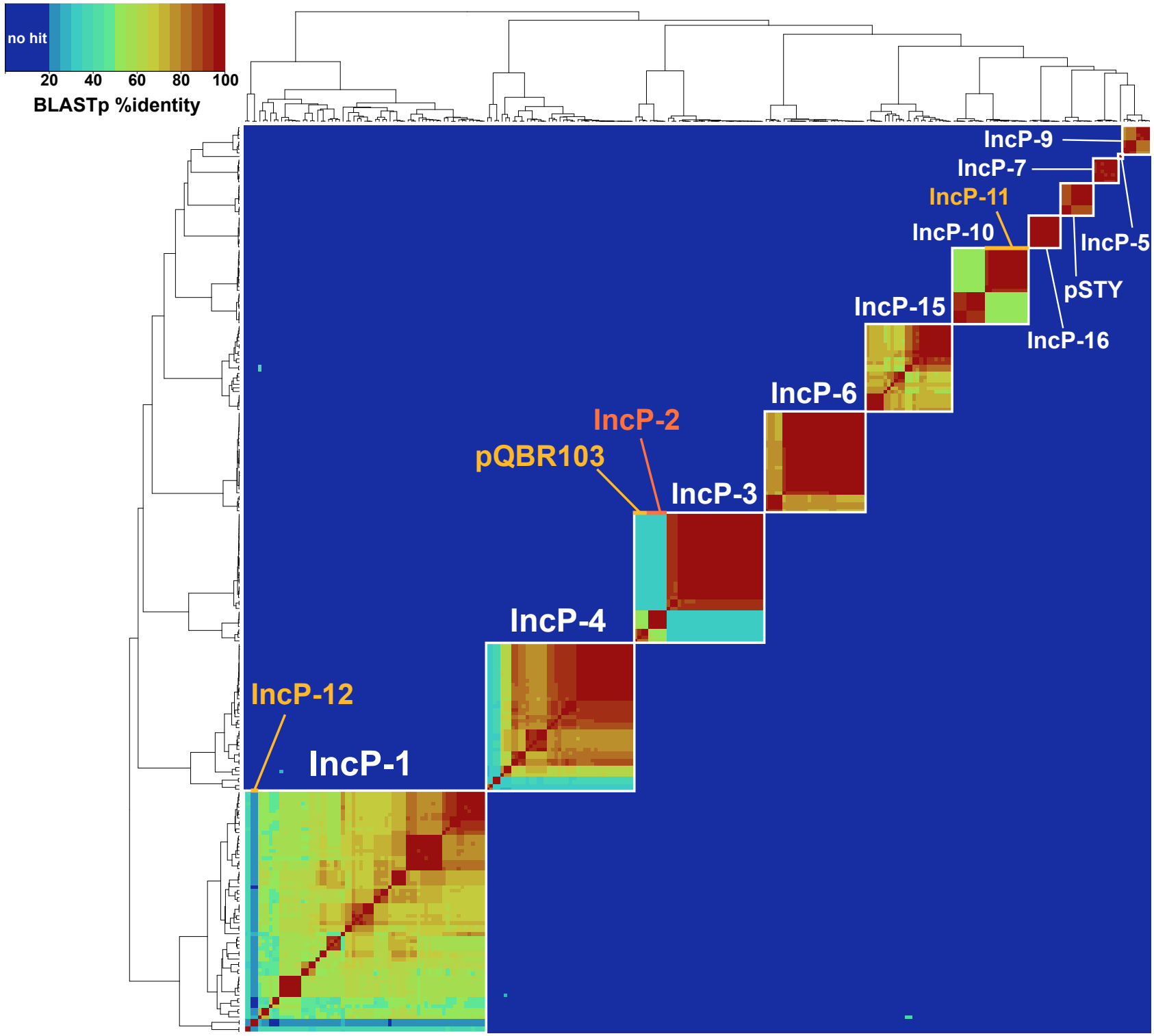
no hit
20
40
60
80
100
BLASTp %identity
IncP-9
IncP-7
IncP-11
IncP-10
IncP-5
pSTY
IncP-15
IncP-16
IncP-6
IncP-2
pQBR103
IncP-3
IncP-4
IncP-12
IncP-1

Fig. 4

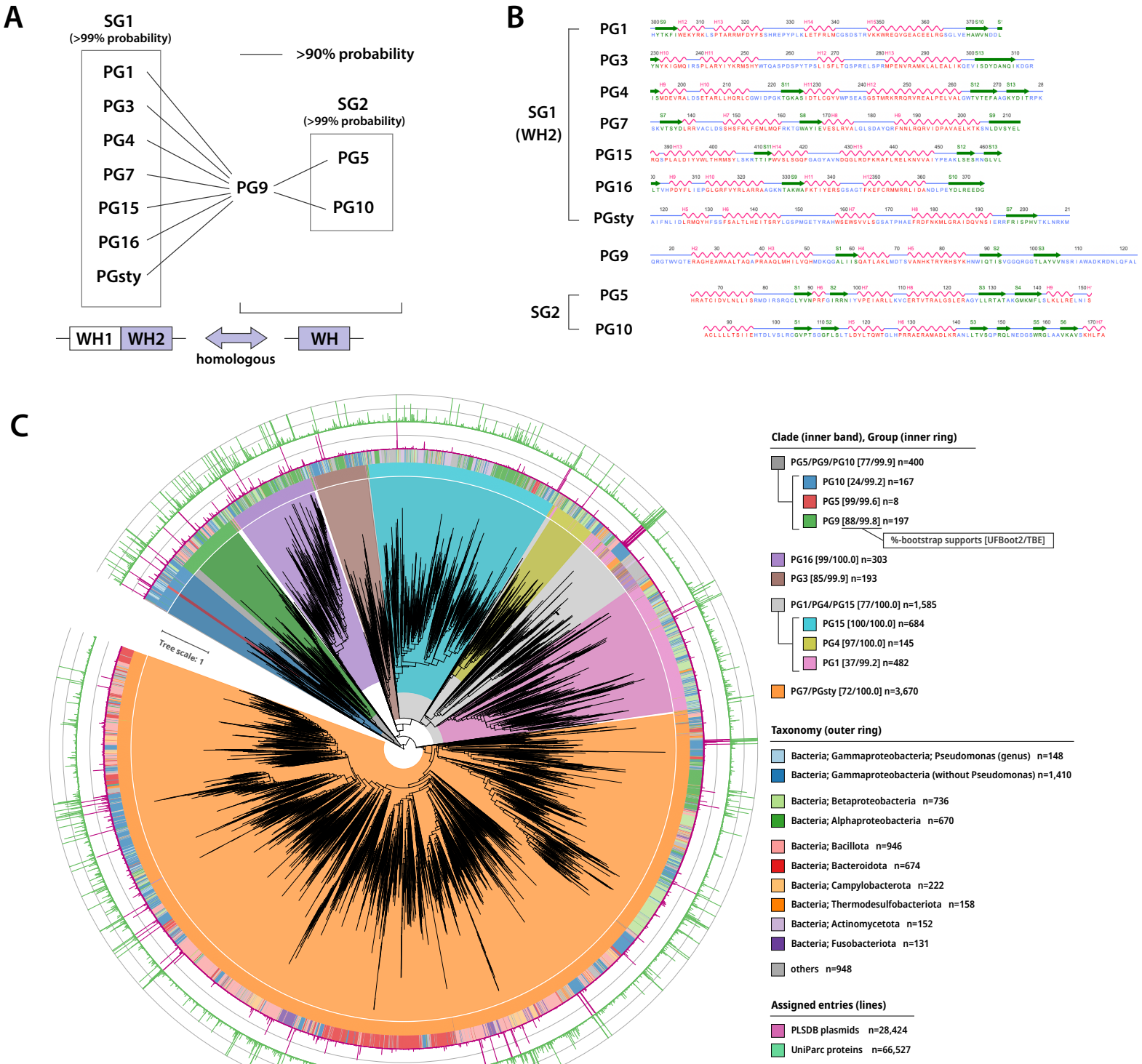
A
SG1
(>99% probability)
PG1
PG3
PG4
PG7
PG15
PG16
PGsty
>90% probability
PG9
SG2
(>99% probability)
PG5
PG10
WH1
WH2
homologous
WH
B
SG1
(WH2)
PG1
PG3
PG4
PG7
PG15
PG16
PGsty
PG9
SG2
PG5
PG10
C
Tree scale: 1
Clade (inner band), Group (inner ring)
PG5/PG9/PG10 [77/99.9] n=400
PG10 [24/99.2] n=167
PG5 [99/99.6] n=8
PG9 [88/99.8] n=197
%-bootstrap supports [UFBoot2/TBE]
PG16 [99/100.0] n=303
PG3 [85/99.9] n=193
PG1/PG4/PG15 [77/100.0] n=1,585
PG15 [100/100.0] n=684
PG4 [97/100.0] n=145
PG1 [37/99.2] n=482
PG7/PGsty [72/100.0] n=3,670
Taxonomy (outer ring)
Bacteria; Gammaproteobacteria; Pseudomonas (genus) n=148
Bacteria; Gammaproteobacteria (without Pseudomonas) n=1,410
Bacteria; Betaproteobacteria n=736
Bacteria; Alphaproteobacteria n=670
Bacteria; Bacillota n=946
Bacteria; Bacteroidota n=674
Bacteria; Campylobacterota n=222
Bacteria; Thermodesulfobacteriota n=158
Bacteria; Actinomycetota n=152
Bacteria; Fusobacteriota n=131
others n=948
Assigned entries (lines)
PLSDB plasmids n=28,424
UniParc proteins n=66,527

Fig. 5