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, Naka-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, Naka-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:+81-53-4781181; 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

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

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

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

114 WH RIPs), exhibiting homology with each other, indicating a shared evolutionary origin. Based
115 on reconstructed phylogeny, we propose a bifurcated evolutionary scenario of WH RIPs; one
116 lineage that includes IncP-5, -9, -10, and -11, and another distinct lineage. (ii) IncP-6, in
117 contrast, belongs to the superfamily of archaeo-eukaryotic primases (AEPs) and represents a
118 distinct evolutionary lineage.

119

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 in DDBJ under the following GenBank accession numbers:
174 LC685025, 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
271 (n=20,331,985) (56) were searched using hmmsearch of HMMER v3.3.2 (57). In the second
272 step, 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

293 **RESULTS AND DISCUSSION**

294 **Overview of plasmid sequences**

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

305

306 **Identification of the RIP gene in each R plasmid**

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

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

316 Initially, no RIP gene was detected in the IncP-5 plasmid Rms163 through nucleotide sequence-
317 based searches. Subsequently, a GC-skew plot of Rms163 was generated (Figure S3), and three
318 nearby CDSs were selected for three-dimensional structure prediction. One CDS showed
319 structural similarity to the replication initiator A family proteins, specifically to Rep19A of pCA-
320 347 with RMSD = 4.2 (Figure 2B). Downstream of this gene, a putative *oriV* region was
321 identified with characteristic repeats, palindromic sequences, a putative DnaA box, and an
322 A+T-rich region (Figure 1). Transformation of *P. aeruginosa* PAO1 with a mini-Rms163A
323 construct containing the RIP gene candidate and *oriV* resulted in a transformation efficiency of
324 4.4×10^1 CFU/µg-DNA. This candidate gene was subsequently named *repP-5A*.

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

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

343 along with the *oriV* region. The constructed mini-RP1-1A successfully transformed *P.*
344 *aeruginosa* at a frequency of 1.3 × 10^2 CFU/µg-DNA (Figure 1), leading to the designation of
345 the CDS as *repP-11A*.

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

353

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

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

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

370 The 99-kb DNA region of pMG26 encoded a MOB_H type relaxase and MPF_G family type IV
371 secretion system (T4SS) (Table 1). To test its conjugative transferability, both filter mating and
372 liquid mating assays were performed between *P. aeruginosa* PAO1808 (donor) carrying pMG26
373 and *P. putida* KT2440 (recipient). The element was successfully transferred from the donor to
374 the recipient strain, with transfer frequencies of 1.4-1.9 × 10^{-5} per donor (filter mating) and
375 0.28-3.4 × 10^{-5} (liquid mating) per donor (Table 3). These results strongly suggest that pMG26
376 functions as an ICE.

377 Additionally, PCR analysis revealed that pMG26 integrated at three different $tRNA_{Lys}$ sites on
378 the KT2440 chromosome (PP_t09, PP_t11 and PP_t13). Notably, the entire DNA region of
379 pMG26 exhibited a high degree of similarity with that of pKLC102 (70) (Figure S5), a member of
380 the PAPI-1 family of ICEs, which are abundant in *P. aeruginosa* (70). Further analyses using

381 MOB-suite indicated that rep_cluster_322 was present in both pMG26 and pKLC102. The
382 putative *oriV* region of pKLC102 (70) was also identified in pMG26.

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

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

393

394 **Transferability of other R-factors**

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

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

404 MOB-suite analysis revealed that the IncP-2 plasmid Rms139 carried the MPF_I family T4SS and
405 MOB_P type relaxase gene (Table 1). In contrast, neither the previously known MOB type
406 relaxase nor the MPF family T4SS components were identified in the IncP-5 plasmid Rms163 or
407 IncP-12 plasmid R716. Instead, both Rms163 and R716 were found to harbour defect in
408 organelle trafficking (Dot) or intracellular multiplication (Icm) systems, which are known as
409 type IVB secretion systems (74). Similarly, the IncP-11 plasmid RP1-1 (GenBank accession
410 number LC700336) lacked the previously known MOB type relaxase and MPF families but
411 possessed putative genes for relaxase (34,618-35,376 nt), type IV coupling protein (32,822-
412 30,702 nt), and proteins involved in mating pair formation.

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

417

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

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

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

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

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

451 The distribution of the 2,357 plasmids across each PIncs group is summarised in Table 1 and
452 Table S4. PIRATE analyses were performed to identify core genes present in 95–100% of the
453 plasmids within each PIncs group (Table 1, S5-1~S5-14), revealing several key features. For
454 instance, the genome sizes of IncP-2, pQBR103-like plasmids, and pSTY-like plasmids were

455 exceptionally large (Table 1, S4, Figure 3A). Additionally, the IncP-1 and IncP-4 plasmids
456 exhibited a wider range of G+C content compared to other groups of plasmids (Figure 3B).
457 Regarding the numbers of core genes, IncP-2, IncP-12, and pQBR103-like plasmids had more
458 core genes than other groups (Table 1, S4, Figure 3C). Furthermore, over 20% of the plasmids
459 in the IncP-2, P-3, P-4, P-6 P-10, and P-11 groups were identified as multi-replicons, meaning
460 that they carried additional RIP genes besides the *repA* gene of PIncs (Table 1, S4, Figure 3D).
461 Importantly, more than 60% of the IncP-2, P-3, P-4, P-6, P-10, and pQBR103-like plasmids
462 harboured ARGs, with several of these groups, IncP-2, P-6, P-10, and pQBR103-like plasmids,
463 containing ARGs of significant clinical importance (Figure 3D).

464

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

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

475

476 **IncP-2 plasmids**

477 Plasmids belonging to the IncP-2 group exhibit significantly larger average DNA sizes of
478 approximately 440 kb. Notably, 293 core genes were identified within this group (Table 1, S4,
479 S5-2), a number significantly higher than that in the other PIncs group. This was around 53% of
480 the average gene numbers (293/549.8) of 87 plasmids in IncP-2 (Table S4). As previously
481 reported (85), the megaplasmid family known as pBT2436 and pRBL16 (14)(15)(36), associated
482 with ARGs, has been recognised as part of the IncP-2 plasmids. Given their large size, specific
483 gene sets in the plasmids are likely crucial for maintaining these plasmids within their hosts.
484 Interestingly, one core gene product exhibited identity with the Tus protein (Table S5-2),
485 known for binding to DNA replication terminus (*ter*) sites of replicons (86). This protein,
486 encoded by rep_cluster_1115 in the MOB-suite database, was found not only in all IncP-2
487 plasmids but also in nine other large-sized plasmids (113–527 kb) and, therefore, was removed
488 from our custom-made RIP reference sequences
489 (https://figshare.com/s/3e3ced8633136259f678).

490

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

492 The IncP-3 group, also known as the IncA/C group, was recently reclassified into two separate
493 Inc groups, IncA and IncC (87). Although RIP genes within IncA and IncC plasmids are highly
494 conserved (>98% identity), the differences in the iteron numbers, spacing, and locations could
495 make them compatible (87). PIRATE analysis revealed that these plasmids share 48 core genes
496 (24.0% of average gene numbers, 48/200.1) (Table 1 and S4), but the accessory genes in the
497 IncP-3 plasmids exhibited greater diversity (Figure 3C). Notably, most IncP-3 (94.5%) plasmids
498 carried at least one ARG (Table 1 and S4), indicating that these plasmids are significant vectors
499 for ARGs.

500

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

502 IncP-4 plasmids, also known as the IncQ group, exhibit a wide range of DNA sizes, from 1.7 to
503 412 kb, with an average of 121 kb (Table 1, Figure 3A). These plasmids are characterised by the
504 presence of *repBAC* genes, which are essential for their single-strand displacement replication
505 (12). However, only two core genes *repA* and *repC* (*repP-4A*) were consistently observed across
506 this group, which was a rather small ratio for their average gene numbers, i.e., 1.5% of 134.1
507 (Table 1, S4, S5-4). This was possibly due to the presence of multi-replicons in the IncP-4
508 group, with only 29.1% plasmids containing a single replicon (Table S4).

509

510 **IncP-5 plasmids**

511 Only one plasmid closely related to Rms163, an IncP-5 plasmid, was identified in the PLSDB
512 (Table 1). This limited representation in the database may not reflect the true prevalence of
513 these plasmids in natural environments. ARGs in Rms163 were located within a Tn*3*-like
514 transposon and a remnant of class 1 integron (Figure S6), with nucleotide sequences showing
515 100% identity to segments of Tn*5393* and class 1 integron previously observed in an IncP-9
516 plasmid, pSN0726-36, isolated from a soil sample in Japan (78). Interestingly, the *repP-5A* gene
517 in Rms163 exhibited 100% identity with the corresponding gene in pPALA50 (GenBank
518 accession number CP111035.1), a gene recently discovered in *P. aeruginosa* isolated from the
519 sputum of a patient with cystic fibrosis (88). Despite the absence of ARGS in pPALA50, the high
520 conservation of other genetic elements between Rms163 and pPALA50 is notable (Figure S6).

521

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

523 IncP-6 plasmids, such as IncP-4, exhibited diverse DNA sizes ranging from 12 to 497 kb (Figure
524 3A). Three core genes were identified, including *repP-6A*, *parC* (involved in plasmid partitioning
525 (89)), and a putative recombinase/transposase gene (Table S5-5). This was a small ratio of
526 average gene numbers, i.e., 5.7% (3/52.7) (Table S4). The archetype plasmid of IncP-6,
527 Rms149, is incompatible with IncU plasmids including pRA3 and pFBAOT6 (90). Although they

528 share the *repP-6A* and *parC* genes, the backbone genes between IncP-6 and IncU plasmids
529 differ (90, 91), suggesting the potential need for further subgroup classification. In this study,
530 the IncP-6 (IncG) and IncU plasmids were separately recognised.

531

532 **IncP-7 plasmids**

533 IncP-7 plasmids shared 10 core genes (5.2% per 191.9 average gene numbers, Table S4),
534 including *repA* (*repP-7A*), and *parWAB* genes, which are involved in plasmid partitioning (92)
535 (Table S5-6). Interestingly, some IncP-7 plasmids, including pWW53 and pND6-1 (93, 94),
536 lacked genes for conjugative transfer, likely explaining why the numbers of core genes were
537 small (Table S4). Among these 23 plasmids, only four carried ARGs, whereas nine others
538 carried catabolic genes, including pCAR1(note that three derivative plasmids has been
539 deposited), pDK1, pWW53, pND6-1, pPHE24 (95), and unnamed plasmids (GenBank accession
540 number NZ_CP071901) (96).

541

542 **IncP-9 plasmids**

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

553

554 **IncP-10 plasmids**

555 All the identified IncP-10 plasmids were hosted by *P. aeruginosa*. More than 50% of them
556 contained ARGs for carbapenem (Table 1). Among the 49 plasmids in the IncP-10 group, the
557 sequence lengths ranged from 29,402 to 106,679 base pairs (bp). Pangenome analysis using
558 PIRATE identified 209 gene families across these plasmids, with 19 core genes present in 95–
559 100% of the plasmids (29.0% per average gene numbers, 19/65.6) (Table S4), including
560 putative genes for plasmid replication and maintenance (Table S5-8).

561

562 **IncP-11 plasmids**

563 IncP-11 plasmids, which are also hosted by *P. aeruginosa* (99), shared 19 core genes, including
564 *repP-11A*, plasmid partition genes, and genes related to conjugative transfer (Table S5-9). The
565 ratio of the core genes per average gene numbers of IncP-10 plasmids was 47.6% (19/65.6,
566 Table S4). Several plasmids were observed in clinical isolates, indicating that IncP-11 plasmids
567 are significant carriers of ARGs in human pathogens. Importantly, pOXA-198 (GenBank
568 accession no. MG958650), previously proposed as an IncP-11 plasmid (100), lacks any genes
569 resembling the *repP-11A*, suggesting that it may not belong to the traditionally recognised
570 IncP-11 group (101).

571

572 **IncP-12 plasmids**

573 Nine plasmids hosted by *P. aeruginosa* were categorised as IncP-12 plasmids (Table 1), with
574 165 core genes, which was a rather large ratio, 74.2% per their average gene numbers
575 (165/222.4, Table S4), although several core genes encoded putative hypothetical proteins
576 (Table S5-10). This indicated that nine plasmids in IncP-12 had high genetic conservation,
577 although none of them other than R716 carried any known ARGs.

578

579 **IncP-15 (PromA)**

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

585

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

587 A total of 24 plasmids were classified as IncP-16 (pSN1216-29), with their sizes ranging from 35
588 to 174 kb. These plasmids shared 17 core genes (28.9% of their average gene numbers,
589 17/58.9) (Table S4) including *repA* and some genes involved in conjugative transfer (Table S5-
590 12). Some of these plasmids, such as pCTX-M15 and pSRC119-A/C, are multi-replicons, which
591 may explain the reduced number of shared genes. Three IncP-16 plasmids carried ARGs.

592

593 **pQBR103-like plasmids**

594 Similar to IncP-2, pQBR103-like plasmids have larger average DNA sizes (~515 kb), with 229
595 core genes identified (Table S5-13). These plasmids are known as pJBCL41-type plasmids and

596 important as they are associated with antimicrobial resistance, especially clinically relevant
597 carbapenem and tigecycline resistance (102, 103).

598

599 **pSTY-like plasmids**

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

605

606 **Phylogenetic analysis of RIPs**

607 To investigate the evolutionary relationships of 15 PIncs, we performed all-against-all BLASTp
608 and analysed the homologous relationships between 251 non-redundant RIP representatives
609 (at 100% sequence identity) (Figure 4). The degree of sequence similarity of RIPs within each
610 PInc was found to vary depending on the PInc. Some PIncs show a high degree of sequence
611 variation, namely IncP-1 (34.05% identity for the most divergent pair), IncP-4 (28.81%), and
612 IncP-15 (56.14%). In contrast, the other PIncs contain less diverse RIP proteins (>73% identity
613 within each PInc) (Table S6). We rarely detected homologous regions with BLASTp between
614 the most sequence pairs of different PIncs, except for some specific combinations of PIncs,
615 such as IncP-2/IncP-3/pQBR103, IncP-1/IncP-12, and IncP-10/IncP-11, which exhibited
616 sequence similarity between different PIncs (Figure 4).

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

627 We then analysed the homologous relationships between the 11 groups. Surprisingly, 10
628 groups exhibited homology with each other, with the exception of PG6, which includes IncP-6
629 plasmids. Based on the highly sensitive HMM–HMM comparisons using HHblits, we identified
630 two supergroups (SG1 and SG2). The groups within each SG shared homology (>99%
631 probability, evalue ≤ 1.5e-22) (Figure 5A). In addition, PG9, which includes IncP-9 plasmids,
632 shared homology (90.1 – 98.4% probability, evalue ≤ 8.8e-05) with all other groups except

633 PG6. The solved and predicted crystal structures of SG1 RIPs, such as RepA from plasmid pPS10
634 (PG7), RepE from plasmid F (PG7), π from plasmid R6K (PG7), and TrfA protein from plasmid
635 RK2 (PG1), contain two winged-helix domains (WH1 and WH2) (105). Although the crystal
636 structures of PG9, PG10, and PG5 RIPs have not been reported, based on the alphafold3
637 predictions, these are predicted to have only one WH domain (Figure 2). We observed that the
638 C-terminal WH domain (WH2) of SG1 RIPs shared homology with the WH domain of PG9 RIPs.
639 WH2 of the RepA protein of pPS10 binds precisely to the iteron 3' end and is the major
640 determinant of RepA sequence specificity (106). We therefore speculated that the conserved
641 WH domain likely plays an important role in replication origin recognition. Here, we named
642 these WH domain-containing RIPs "WH RIPs", which constitute a "WH RIP superfamily".

643 The remaining PG6 exhibited no homology with the other groups (<1% probability by HHblits).
644 To characterise the evolutionary origin of PG6, we performed a HHblits search against Pfam
645 v35.0. The RIPs of PG6 show significant homology to PF03090 (Replicase family; with 98.4%
646 probability by HHblits), including the ColE2 Rep protein. The ColE2 Rep protein belongs to the
647 superfamily of AEPs (107) and is the only plasmid-specific trans-acting factor required for
648 initiation of plasmid replication (108). The RIPs of IncP-6 plasmids (9, 89) have a PriCT domain,
649 which may form a compact module and play a central role in DNA unwinding. The C termini of
650 some AEPs suggest that it probably plays a similar role in these proteins. The RIPs of PG6
651 plasmids have domain compositions different from those of the other groups. We therefore
652 concluded that RIPs of PG6 have different evolutionary origins with the other groups.

653 We reconstructed the phylogeny of the WH RIPs using the identified homologous region of the
654 WH domain. The homologous region showed similar secondary structure compositions
655 including multiple alpha helixes and beta turns among the 10 groups of the WH RIPs (Figure
656 5B). A maximum-likelihood phylogenetic tree of the WH RIPs was reconstructed, which
657 comprised 6,194 clustering representatives (60% identity) associated with 28,424 PLSDB
658 plasmids and 66,527 Uniparc proteins (Figure 5C). The tree is subdivided largely according to
659 the 10 groups with two special notes. First, PG7 (including IncP-7 plasmids) and PGsty
660 (including pSTY plasmids) were merged and formed one large group because these are being
661 composed of similar sequences. Second, PG5 (including IncP-5 plasmids) formed a subclade
662 within the clade of PG10. We defined clades considering the group affiliations and branch
663 support values. Most clades (i.e., PG5, PG9, PG16, PG3, PG15, PG4, and PG7/PGsty) were
664 identified with two types of high support values (UFBoot2: 72–100%, TBE: 99–100%). The
665 remaining (PG1 and PG10) clades were not highly supported by UFBoot2 but were supported
666 by TBE (>99%). Considering the larger structure, we defined a superclade PG1/PG4/PG15,
667 containing the PG1, PG4, and PG15 clades, and a superclade PG5/PG9/PG10, containing the
668 PG5, PG9, and PG10 clades, both of which were supported with high support values (UFBoot2:
669 77%, TBE: 99.9–100%). Notably, the branch dividing the PG5/PG9/PG10 superclade and the
670 remaining tree constitutes the well-supported division of SG1, which is composed of two WH
671 domains, and SG2/PG9, which is composed of one WH domain. We thus speculated that this
672 division is possibly the most ancestral split in the tree.

673 The tree includes diverse sequences not derived from PLSDB plasmids, indicating that
674 numerous uncharacterised plasmids and incompatibility groups exist. In total, 75.4% (n =
675 4,669) of tree leaves were not associated with PLSDB plasmids (Figure 5C). Conversely, a
676 substantial proportion of PLSDB plasmids (47.4%; n = 28,424) was associated with the
677 remaining leaves of the tree (n = 1,525), indicating that these plasmids are likely replicated by
678 WH RIPs. Considering plasmids with alternative replication mechanisms that employ disparate
679 types of RIPs, such as the superfamily of AEPs (e.g., IncP-6) and rolling-circle replication (e.g.,
680 Pfam PF01719; Rep_OBD), this fraction is notably high, almost reaching 29,662 PLSDB plasmids
681 identified with plasmid types using PlasmidFinder v2.1
682 (https://figshare.com/s/915a336484ee084ffbea; 59,895 plasmid list). Notably, 9,666 plasmids
683 lacking PlasmidFinder types were associated with the tree. In other words, 8,860 additional
684 plasmids were associated with the tree even without counting the 2,357 Pincs plasmids. This
685 indicates that the tree can serve as a reference for the characterisation of plasmids, facilitating
686 the identification and classification of RIP sequences. Furthermore, among the 1,525 leaves
687 associated with PLSDB plasmids, 51 leaves are associated with more than 100 plasmids (102–
688 6,872), accumulating 32,713 RIPs (representing 81.0% of the total 40,371 RIPs), which suggests
689 a biased knowledge of plasmids containing similar RIPs. Subsequently, 9,141 plasmids (32.2%)
690 were identified with multiple RIPs, suggesting that multi-replicons are not a special case of
691 plasmids.

692 Each clade of the tree contains diverse taxonomic lineages, suggesting potential host
693 transitions within the clade (Figure 5C). In contrast, some taxa are enriched in specific clades.
694 For instance, the phylum Actinomycetota is enriched in PG15 (84.2%), whereas the phyla
695 Bacillota, Bacteroidota, Campylobacterota, and Fusobacteriota are enriched in PG7/PGsty
696 (90.0, 93.0, 95.0, 100%, respectively). Notably, the latter lineages are enriched in the left-wing
697 subclade of the PG7/PGsty. Although this enrichment might be associated with some specific
698 environment, further study is required to describe the detailed association between clades,
699 lineages, and environments.

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

707

708 **CONCLUSION**

709 This study provides valuable insights into the classification and genetic characterisation of
710 *Pseudomonas* plasmids. We successfully determined the complete whole nucleotide
711 sequences of several previously uncharacterised plasmids, including those in the

712 *Pseudomonas*, namely the IncP-5, IncP-11, and IncP-12 groups, and reclassified plasmids that
713 were previously thought to belong to 'IncP-13' plasmids as an ICE. Furthermore, we identified
714 RIP genes associated with these plasmids, along with the IncP-2 plasmid, contributing to a
715 more refined understanding of their replication mechanisms.

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

722 Our findings underscore the strengths of our classification system, which is grounded in
723 experimental data and aligns closely with established knowledge. In contrast, the recently
724 launched IMG/Plasmid database, despite its comprehensiveness in terms of including a vast
725 number of plasmid sequences (>700,000 sequences) (https://img.jgi.doe.gov/cgi-
726 bin/plasmid/main.cgi)(109), employs plasmid taxonomic units (PTUs) based solely on average
727 nucleotide sequence identity (110). Although this method is thorough in scope, it fails to
728 capture the distinct characteristics of individual plasmid groups that have been documented
729 through years of research as they are affected by long-sized accessory genes in plasmids. The
730 experimental identification of genes and DNA regions involved in plasmid replication remains
731 crucial for a deeper understanding of plasmid biology, including their host range, coevolution
732 with bacterial hosts, and potential applications in biotechnology. These findings are
733 particularly relevant for understanding the spread of ARG-associated plasmids in clinical and
734 environmental *Pseudomonas* isolates, especially in the context of multidrug-resistant *P.*
735 *aeruginosa* strains.

736 The reconstructed phylogenetic tree of WH RIPs will provide a fundamental basis for plasmid
737 detection and classification based on RIPs, which are essential proteins for understanding
738 plasmid replication mechanisms. Further research should focus on integrating experimental
739 data with large-scale sequence analyses to reconcile discrepancies between traditional and
740 modern classification systems. This will enhance our ability to track and manage antibiotic
741 resistance in various environments, ultimately contributing to better public health outcomes.

742

743 **DATA AVAILABILITY**

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

746

747 **Supplementary Data statement**
748 Supplementary Data are available at NAR online.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGEMENTS

We are grateful to Prof. Dr. S. Iyobe of the Gumma University School of Medicine and Prof. Dr. 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

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

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

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

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

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

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

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

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

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

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

805 11. Yano,H., Shintani,M., Tomita,M., Suzuki,H. and Oshima,T. (2019) Reconsidering plasmid
806 maintenance factors for computational plasmid design. *Comput. Struct. Biotechnol. J.*,
807 **17**, 70–81.

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

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

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

818 15. Jiang,X., Yin,Z., Yuan,M., Cheng,Q., Hu,L., Xu,Y., Yang,W., Yang,H., Zhao,Y., Zhao,X., *et al.*
819 (2020) Plasmids of novel incompatibility group IncpRBL16 from Pseudomonas species.
820 *J. Antimicrob. Chemother.*, **75**, 2093–2100.

821 16. Urbanowicz,P., Bitar,I., Izdebski,R., Baraniak,A., Literacka,E., Hrabák,J. and Gniadkowski,M.
822 (2021) Epidemic Territorial Spread of IncP-2-Type VIM-2 Carbapenemase-Encoding
823 Megaplasmids in Nosocomial Pseudomonas aeruginosa Populations. *Antimicrob.*
824 *Agents Chemother.*, **65**.

825 17. Jacoby,G.A. (1986) Resistance Plasmids of Pseudomonas. *The Biology of Pseudomonas*,
826 10.1016/b978-0-12-307210-8.50013-0.

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

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

833 20. Summers,A.O. and Jacoby,G.A. (1978) Plasmid-determined resistance to boron and
834 chromium compounds in Pseudomonas aeruginosa. *Antimicrob. Agents Chemother.*,
835 **13**, 637–640.

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

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

842 23. Jacoby,G.A. (1980) PLASMID DETERMINED RESISTANCE TO CARBENICILLIN AND
843 GENTAMICIN IN Pseudomonas aeruginosa. *Plasmids and Transposons*, 10.1016/b978-
844 0-12-675550-3.50011-3.

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

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

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

851 27. Schmartz,G.P., Hartung,A., Hirsch,P., Kern,F., Fehlmann,T., Müller,R. and Keller,A. (2022)
852 PLSDB: advancing a comprehensive database of bacterial plasmids. *Nucleic Acids Res.*,
853 **50**, D273–D278.

854 28. Galata,V., Fehlmann,T., Backes,C. and Keller,A. (2018) PLSDB: a resource of complete
855 bacterial plasmids. *Nucleic Acids Res.*, 10.1093/nar/gky1050.

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

858 30. Shintani,M., Yoshida,T., Habe,H., Omori,T. and Nojiri,H. (2005) Large plasmid pCAR2 and
859 class II transposon Tn4676 are functional mobile genetic elements to distribute the
860 carbazole/dioxin-degradative car gene cluster in different bacteria. *Appl. Microbiol.*
861 *Biotechnol.*, **67**, 370–382.

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

864 32. plasmidSPAdes – Center for Algorithmic Biotechnology.

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

869 34. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G.,
870 Durbin,R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence
871 Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

872 35. Shen,W., Le,S., Li,Y. and Hu,F. (2016) SeqKit: A Cross-Platform and Ultrafast Toolkit for
873 FASTA/Q File Manipulation. *PLoS One*, **11**, e0163962.

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

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

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

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

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

889 41. Feldgarden,M., Brover,V., Gonzalez-Escalona,N., Frye,J.G., Haendiges,J., Haft,D.H.,
890 Hoffmann,M., Pettengill,J.B., Prasad,A.B., Tillman,G.E., *et al.* (2021) AMRFinderPlus

891 and the Reference Gene Catalog facilitate examination of the genomic links among
892 antimicrobial resistance, stress response, and virulence. *Sci. Rep.*, **11**, 12728.

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

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

896 44. Solovyev,V. (2011) V. Solovyev, A Salamov (2011) Automatic Annotation of Microbial
897 Genomes and Metagenomic Sequences. In Metagenomics and its Applications in
898 Agriculture, Biomedicine and Environmental Studies (Ed. R.W. Li), Nova Science
899 Publishers, p.61-78.

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

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

906 47. Swan,M.K., Bastia,D. and Davies,C. (2006) Crystal structure of π initiator protein–iteron
907 complex of plasmid R6K: Implications for initiation of plasmid DNA replication.
908 *Proceedings of the National Academy of Sciences*, **103**, 18481–18486.

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

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

914 50. Carattoli,A. and Hasman,H. (2020) PlasmidFinder and in silico pMLST: Identification and
915 typing of Plasmid replicons in whole-genome sequencing (WGS). *Methods Mol. Biol.*,
916 **2075**, 285–294.

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

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

922 53. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J.
923 (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search
924 programs. *Nucleic Acids Res.*, **25**, 3389–3402.

925 54. Fassler,J. and Cooper,P. (2011) BLAST Glossary National Center for Biotechnology
926 Information (US).

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

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

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

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

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

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

940 61. Minh,B.Q., Schmidt,H.A., Chernomor,O., Schrempf,D., Woodhams,M.D., von Haeseler,A.
941 and Lanfear,R. (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
942 Inference in the Genomic Era. *Mol. Biol. Evol.*, **37**, 1530–1534.

943 62. Kalyaanamoorthy,S., Minh,B.Q., Wong,T.K.F., von Haeseler,A. and Jermiin,L.S. (2017)
944 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods*,
945 **14**, 587–589.

946 63. Hoang,D.T., Chernomor,O., von Haeseler,A., Minh,B.Q. and Vinh,L.S. (2018) UFBoot2:
947 Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.*, **35**, 518–522.

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

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

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

955 67. Kawakami,Y., Mikoshiba,F., Nagasaki,S., Matsumoto,H. and Tazaki,T. (1972) Prevalence of
956 Pseudomonas aeruginosa strains possessing R factor in a hospital. *J. Antibiot.* , **25**,
957 607–609.

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

961 69. Kawalek,A., Kotecka,K., Modrzejewska,M., Gawor,J., Jagura-Burdzy,G. and Bartosik,A.A.
962 (2020) Genome sequence of Pseudomonas aeruginosa PAO1161, a PAO1 derivative
963 with the ICEPae1161 integrative and conjugative element. *BMC Genomics*, **21**, 14.

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

966 71. Johnson,C.M. and Grossman,A.D. (2015) Integrative and Conjugative Elements (ICEs): What
967 They Do and How They Work. *Annu. Rev. Genet.*, **49**, 577–601.

968 72. Bradley,D.E., Taylor,D.E. and Cohen,D.R. (1980) Specification of surface mating systems
969 among conjugative drug resistance plasmids in Escherichia coli K-12. *J. Bacteriol.*, **143**,
970 1466–1470.

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

974 74. Kitao,T., Kubori,T. and Nagai,H. (2022) Recent advances in structural studies of the
975 Legionella pneumophila Dot/Icm type IV secretion system. *Microbiol. Immunol.*, **66**,
976 67–74.

977 75. Yanagiya,K., Maejima,Y., Nakata,H., Tokuda,M., Moriuchi,R., Dohra,H., Inoue,K.,
978 Ohkuma,M., Kimbara,K. and Shintani,M. (2018) Novel Self-Transmissible and Broad-
979 Host-Range Plasmids Exogenously Captured From Anaerobic Granules or Cow Manure.
980 *Front. Microbiol.*, **9**, 2602.

981 76. Tokuda,M., Suzuki,H., Yanagiya,K., Yuki,M., Inoue,K., Ohkuma,M., Kimbara,K. and
982 Shintani,M. (2020) Determination of Plasmid pSN1216-29 Host Range and the
983 Similarity in Oligonucleotide Composition Between Plasmid and Host Chromosomes.
984 *Front. Microbiol.*, **11**, 1187.

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

988 78. Hayakawa,M., Tokuda,M., Kaneko,K., Nakamichi,K., Yamamoto,Y., Kamijo,T., Umeki,H.,
989 Chiba,R., Yamada,R., Mori,M., *et al.* (2022) Hitherto-Unnoticed Self-Transmissible
990 Plasmids Widely Distributed among Different Environments in Japan. *Appl. Environ.*
991 *Microbiol.*

992 79. Lilley,A.K. and Bailey,M.J. (1997) Impact of Plasmid pQBR103 Acquisition and Carriage on
993 the Phytosphere Fitness of Pseudomonas fluorescens SBW25: Burden and Benefit.
994 *Appl. Environ. Microbiol.*, **63**, 1584–1587.

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

999 81. Lv,L., Wan,M., Wang,C., Gao,X., Yang,Q., Partridge,S.R., Wang,Y., Zong,Z., Doi,Y., Shen,J., *et*
1000 *al.* (2020) Emergence of a Plasmid-Encoded Resistance-Nodulation-Division Efflux
1001 Pump Conferring Resistance to Multiple Drugs, Including Tigecycline, in Klebsiella
1002 pneumoniae. *MBio*, **11**.

1003 82. Köhler,K.A.K., Rückert,C., Schatschneider,S., Vorhölter,F.-J., Szczepanowski,R., Blank,L.M.,
1004 Niehaus,K., Goesmann,A., Pühler,A., Kalinowski,J., *et al.* (2013) Complete genome
1005 sequence of Pseudomonas sp. strain VLB120 a solvent tolerant, styrene degrading
1006 bacterium, isolated from forest soil. *J. Biotechnol.*, **168**, 729–730.

1007 83. Chen,F., Wang,P., Yin,Z., Yang,H., Hu,L., Yu,T., Jing,Y., Guan,J., Wu,J. and Zhou,D. (2022)
1008 VIM-encoding IncpSTY plasmids and chromosome-borne integrative and mobilizable
1009 elements (IMEs) and integrative and conjugative elements (ICEs) in Pseudomonas.
1010 *Ann. Clin. Microbiol. Antimicrob.*, **21**, 10.

1011 84. Pachulec,E. and van der Does,C. (2010) Conjugative plasmids of Neisseria gonorrhoeae.
1012 *PLoS One*, **5**, e9962.

1013 85. Shintani,M., Suzuki,H., Nojiri,H. and Suzuki,M. (2022) Precise classification of antimicrobial
1014 resistance-associated IncP-2 megaplasmids for molecular epidemiological studies on
1015 Pseudomonas species. *J. Antimicrob. Chemother.*, 10.1093/jac/dkac006.

1016 86. Hidaka,M., Kobayashi,T., Takenaka,S., Takeya,H. and Horiuchi,T. (1989) Purification of a
1017 DNA replication terminus (ter) site-binding protein in Escherichia coli and identification
1018 of the structural gene. *J. Biol. Chem.*, **264**, 21031–21037.

1019 87. Ambrose,S.J., Harmer,C.J. and Hall,R.M. (2018) Compatibility and entry exclusion of IncA
1020 and IncC plasmids revisited: IncA and IncC plasmids are compatible. *Plasmid*, **96–97**, 7–
1021 12.

1022 88. Robinson,L.A., Collins,A.C.Z., Murphy,R.A., Davies,J.C. and Allsopp,L.P. (2023) Diversity and
1023 prevalence of type VI secretion system effectors in clinical Pseudomonas aeruginosa
1024 isolates. *Front. Microbiol.*, **13**.

1025 89. Haines,A.S., Jones,K., Cheung,M. and Thomas,C.M. (2005) The IncP-6 plasmid Rms149
1026 consists of a small mobilizable backbone with multiple large insertions. *J. Bacteriol.*,
1027 **187**, 4728–4738.

1028 90. Kulinska,A., Czeredys,M., Hayes,F. and Jagura-Burdzy,G. (2008) Genomic and functional
1029 characterization of the modular broad-host-range RA3 plasmid, the archetype of the
1030 IncU group. *Appl. Environ. Microbiol.*, **74**, 4119–4132.

1031 91. Rhodes,G., Parkhill,J., Bird,C., Ambrose,K., Jones,M.C., Huys,G., Swings,J. and Pickup,R.W.
1032 (2004) Complete nucleotide sequence of the conjugative tetracycline resistance
1033 plasmid pFBAOT6, a member of a group of IncU plasmids with global ubiquity. *Appl.*
1034 *Environ. Microbiol.*, **70**, 7497–7510.

1035 92. Shintani,M., Yano,H., Habe,H., Omori,T., Yamane,H., Tsuda,M. and Nojiri,H. (2006)
1036 Characterization of the replication, maintenance, and transfer features of the IncP-7

1037 plasmid pCAR1, which carries genes involved in carbazole and dioxin degradation.
1038 *Appl. Environ. Microbiol.*, **72**, 3206–3216.

1039 93. Yano,H., Garruto,C.E., Sota,M., Ohtsubo,Y., Nagata,Y., Zylstra,G.J., Williams,P.A. and
1040 Tsuda,M. (2007) Complete sequence determination combined with analysis of
1041 transposition/site-specific recombination events to explain genetic organization of
1042 IncP-7 TOL plasmid pWW53 and related mobile genetic elements. *J. Mol. Biol.*, **369**,
1043 11–26.

1044 94. Yano,H., Miyakoshi,M., Ohshima,K., Tabata,M., Nagata,Y., Hattori,M. and Tsuda,M. (2010)
1045 Complete nucleotide sequence of TOL plasmid pDK1 provides evidence for
1046 evolutionary history of IncP-7 catabolic plasmids. *J. Bacteriol.*, **192**, 4337–4347.

1047 95. Heinaru,E., Truu,J., Stottmeister,U. and Heinaru,A. (2000) Three types of phenol and p-
1048 cresol catabolism in phenol- and p-cresol-degrading bacteria isolated from river water
1049 continuously polluted with phenolic compounds. *FEMS Microbiol. Ecol.*, **31**, 195–205.

1050 96. Lopez-Echartea,E., Suman,J., Smrhova,T., Ridl,J., Pajer,P., Strejcek,M. and Uhlik,O. (2021)
1051 Genomic analysis of dibenzofuran-degrading Pseudomonas veronii strain Pvy reveals
1052 its biodegradative versatility. *G3* , **11**.

1053 97. Suenaga,H., Fujihara,H., Kimura,N., Hirose,J., Watanabe,T., Futagami,T., Goto,M.,
1054 Shimodaira,J. and Furukawa,K. (2017) Insights into the genomic plasticity of
1055 Pseudomonas putida KF715, a strain with unique biphenyl-utilizing activity and
1056 genome instability properties. *Environ. Microbiol. Rep.*, **9**, 589–598.

1057 98. Sevastsyanovich,Y.R., Krasowiak,R., Bingle,L.E., Haines,A.S., Sokolov,S.L., Kosheleva,I.A.,
1058 Leuchuk,A.A., Titok,M.A., Smalla,K. and Thomas,C.M. (2008) Diversity of IncP-9
1059 plasmids of Pseudomonas. *Microbiology*, **154**, 2929–2941.

1060 99. Huang,W., Hamouche,J.E., Wang,G., Smith,M., Yin,C., Dhand,A., Dimitrova,N. and
1061 Fallon,J.T. (2020) Integrated Genome-Wide Analysis of an Isogenic Pair of
1062 Pseudomonas aeruginosa Clinical Isolates with Differential Antimicrobial Resistance to
1063 Ceftolozane/Tazobactam, Ceftazidime/Avibactam, and Piperacillin/Tazobactam. *Int. J.*
1064 *Mol. Sci.*, **21**.

1065 100. Bonnin,R.A., Bogaerts,P., Girlich,D., Huang,T.-D., Dortet,L., Glupczynski,Y. and Naas,T.
1066 (2018) Molecular Characterization of OXA-198 Carbapenemase-Producing
1067 Pseudomonas aeruginosa Clinical Isolates. *Antimicrob. Agents Chemother.*, **62**.

1068 101. Shintani,M., Suzuki,H., Nojiri,H. and Suzuki,M. (2023) Reconsideration of the previously
1069 classified incompatibility groups of plasmids, IncP-1 and IncP-11. *Environ. Microbiol.*,
1070 10.1111/1462-2920.16345.

1071 102. Botelho,J., Lood,C., Partridge,S.R., van Noort,V., Lavigne,R., Grosso,F. and Peixe,L. (2019)
1072 Combining sequencing approaches to fully resolve a carbapenemase-encoding
1073 megaplasmid in a *Pseudomonas shirazica* clinical strain. *Emerg. Microbes Infect.*, **8**,
1074 1186–1194.

1075 103. Zhu,J., Lv,J., Zhu,Z., Wang,T., Xie,X., Zhang,H., Chen,L. and Du,H. (2023) Identification of
1076 TMexCD-TOprJ-producing carbapenem-resistant Gram-negative bacteria from hospital
1077 sewage. *Drug Resist. Updat.*, **70**, 100989.

1078 104. Chen,M., Cai,H., Li,Y., Wang,N., Zhang,P., Hua,X., Yu,Y. and Sun,R. (2022) Plasmid-Borne
1079 AFM Alleles in Pseudomonas aeruginosa Clinical Isolates from China. *Microbiol. Spectr.*,
1080 **10**.

1081 105. Wegrzyn,K., Zabrocka,E., Bury,K., Tomiczek,B., Wieczor,M., Czub,J., Uciechowska,U.,
1082 Moreno-Del Alamo,M., Walkow,U., Grochowina,I., *et al.* (2021) Defining a novel
1083 domain that provides an essential contribution to site-specific interaction of Rep
1084 protein with DNA. *Nucleic Acids Res.*, 10.1093/nar/gkab113.

1085 106. Giraldo,R. and Fernandez-Tresguerres,M.E. (2004) Twenty years of the pPS10 replicon:
1086 insights on the molecular mechanism for the activation of DNA replication in iteron-
1087 containing bacterial plasmids. *Plasmid*, **52**, 69–83.

1088 107. Iyer,L.M., Koonin,E.V., Leipe,D.D. and Aravind,L. (2005) Origin and evolution of the
1089 archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural
1090 insights and new members. *Nucleic Acids Res.*, **33**, 3875–3896.

1091 108. Itou,H., Yagura,M., Shirakihara,Y. and Itoh,T. (2015) Structural basis for replication origin
1092 unwinding by an initiator primase of plasmid ColE2-P9: duplex DNA unwinding by a
1093 single protein. *J. Biol. Chem.*, **290**, 3601–3611.

1094 109. Camargo,A.P., Roux,S., Schulz,F., Babinski,M., Xu,Y., Hu,B., Chain,P.S.G., Nayfach,S. and
1095 Kyrpides,N.C. (2023) Identification of mobile genetic elements with geNomad. *Nat.*
1096 *Biotechnol.*, 10.1038/s41587-023-01953-y.

1097 110. Redondo-Salvo,S., Bartomeus-Peñalver,R., Vielva,L., Tagg,K.A., Webb,H.E., Fernández-
1098 López,R. and de la Cruz,F. (2021) COPLA, a taxonomic classifier of plasmids. *BMC*
1099 *Bioinformatics*, **22**, 390.

1100 111. Robertson,J. and Nash,J.H.E. (2018) MOB-suite: software tools for clustering,
1101 reconstruction and typing of plasmids from draft assemblies. *Microbial genomics*.

1102 112. Holloway,B.W. (1955) Genetic recombination in Pseudomonas aeruginosa. *J. Gen.*
1103 *Microbiol.*, **13**, 572–581.

1104 113. Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P., Hickey,M.J.,
1105 Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M., *et al.* (2000) Complete genome
1106 sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature*, **406**,
1107 959–964.

1108 114. Bagdasarian,M., Lurz,R., Rückert,B., Franklin,F.C., Bagdasarian,M.M., Frey,J. and
1109 Timmis,K.N. (1981) Specific-purpose plasmid cloning vectors. II. Broad host range, high
1110 copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in
1111 Pseudomonas. *Gene*, **16**, 237–247.

1112 115. Nelson,K.E., Weinel,C., Paulsen,I.T., Dodson,R.J., Hilbert,H., Martins dos Santos,V.A.,
1113 Fouts,D.E., Gill,S.R., Pop,M., Holmes,M., *et al.* (2002) Complete genome sequence and
1114 comparative analysis of the metabolically versatile Pseudomonas putida KT2440.
1115 *Environ. Microbiol.*, **4**, 799–808.

1116 116. Holloway,B.W. and Richmond,M.H. (1973) R-Factors used for genetic studies in strains of
1117 Pseudomonas aeruginosa and their origin. *Genet. Res.* , **21**, 103–105.

1118 117. Grinsted,J., Saunders,J.R., Ingram,L.C., Sykes,R.B. and Richmond,M.H. (1972) Properties of
1119 a R factor which originated in Pseudomonas aeruginosa 1822. *J. Bacteriol.*, **110**, 529–
1120 537.

1121 **TABLES**

1122 **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]

1123 [a]Classification of MOB classes and MPF types was identified by the MOB-suite (111) with criteria of >=90% identity and >=60% coverage. "-"
1124 indicates that the genes involved in conjugation have not been detected, whereas "NA" indicates that the classes and types of MOB or MPF were
1125 not identified although the plasmid was self-transmissible. [b]Core genes of each Inc group plasmid were identified by PIRATE analyses. [c]Partial DNA
1126 sequences are available. [d]pMG26 was shown to be not a plasmid but an integrative and conjugative element in this study.

1127 **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	(112, 113)
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	(114, 115)
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	(116, 117)
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

1128

1129 **Table 3. Conjugation frequency of each plasmid assessed with *Pseudomonas aeruginosa***
1130 **strains PAO1 and KT2440 as recipients, using both filter (solid) mating and liquid mating**
1131 **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

1134 **Figures**

1135 **Figure 1. Genetic structure of mini-replicon of IncP-2 (Rms139), IncP-5 (Rms163), IncP-10**
1136 **(pAP546), IncP-11 (RP1-1), and IncP-12 (R716) plasmids.**
1137 (A) The mini-replicon of each plasmid was constructed by using DNA regions shown in a red
1138 arrow (replication initiation protein RIP gene) with its promoter (grey box) and a grey solid line
1139 (*oriV*). Putative iterons, palindromes, and A+T-rich regions are shown by short red arrows,
1140 short yellow arrows, and blue boxes below the *oriV* region, respectively. Putative DnaA boxes
1141 are shown by black arrowheads in the *oriV* region. (B) Consensus sequences of putative iterons
1142 are shown. (C) Transformation efficiency of *P. aeruginosa* PAO1 with the mini-replicons are
1143 shown.

1144 **Figure 2. Predicted three-dimensional structure of RIP of plasmids in PIncs.**
1145 (A) Predicted structures of RIPs of IncP-1 (RepP-1A (=TrfA1) of RK2), IncP-2 (RepP-2A of
1146 Rms139), IncP-3 (RepP-3A of pRA1), IncP-4 (RepP-4A (=RepC) of RSF1010), IncP-7 (RepP-7A of
1147 pCAR1), IncP-9 (RepP-9A of pWW0), IncP-12 (RepP-12A of R716), IncP-15 (RepP-15A of
1148 pSN1104-11), IncP-16 (RepP-16A of pSB1216-29), pQBR103 ($Rep_{pQBR103}$), and pSTY (Rep_{pSTY}).
1149 The crystal structure of Pi protein of IncX plasmid, R6K, is shown as reference, of which
1150 structure is elucidated by X-ray crystallography. (B) Predicted structures of RIPs of IncP-5
1151 (RepP-5A of Rms163), IncP-10 (RepP-10A of pPAB546), and IncP-11 (RepP-11A of RP1-1). The
1152 crystal structure of Rep19A protein of *Staphylococcus* plasmid, pCA-347, is shown as reference,
1153 of which structure is elucidated by X-ray crystallography.

1154 **Figure 3. Features of plasmids of PIncs in the PLSDB.**

1155 Distribution of (A) plasmid sizes (kb) and (B) G+C contents (%), and (C) ratio of core and
1156 accessory genes of each PIncs plasmids in the PLSDB. (D) Relationship of each PIncs plasmids to
1157 other plasmid replicons and ARGs in the PLSDB. A sankey diagram was generated using the
1158 SankeyWidget library in Python through in-house programming.

1159 **Figure 4. Homologous relationships between 251 non-redundant RIPs of 15 PIncs.**

1160 %-identities of all-against-all BLASTp among the 251 non-redundant RIPs were visualised. The
1161 heatmap and the dendrograms were generated using R v3.6.1 and R packages dendextend and
1162 gplots via CPAP v0.1.0 (https://github.com/yosuken/CPAP) with a option ("--clust-method
1163 complete") to perform clustering with complete-linkage method.

1164 **Figure 5. Evolutionary relationships and phylogeny of WH RIPs.**

1165 (A) Remote homology among 10 WH RIP groups, identified using HHblits. Two supergroups
1166 (SG1 and SG2) were defined according to the high probability (>99%) predicted within each
1167 supergroup. Relationships with more than 90% probability were indicated by lines. The WH2
1168 domain of SG1 is homologous to the WH domain of SG2 and PG9. The domain lengths shown
1169 are arbitrary. (B) The secondary structures of representatives of the 10 WH RIP groups. The
1170 representatives were selected from the 3D structure-resolved sequences shown in Figure 2.
1171 Namely, PG1 (RepP-1A (=TrfA1) of RK2), PG3 (RepP-3A of pRA1), PG4 (RepP-4A (=RepC) of

1172 RSF1010), PG7 (RepP-7A of pCAR1), PG15 (RepP-15A of pSN1104-11), PG16 (RepP-16A of
1173 pSB1216-29), PGsty (ReppSTY), PG9 (RepP-9A of pWW0), PG5 (RepP-5A of Rms163), and PG10
1174 (RepP-10A of pPAB546) were selected. The secondary structures were predicted and visualised
1175 using iCn3D based on the 3D structures. (C) The ML phylogenetic tree of WH RIPs (6,194
1176 leaves) calculated based on the conserved region (98 positions in the alignment) of the WH
1177 domain. The tree was rooted by the superclade of PG5/PG9/PG10. The clade and superclade
1178 names, bootstrap values, and number of leaves are shown in the legend. The consensus
1179 taxonomic affiliation of each leaf is shown in the outer ring. The number of leaves for each
1180 taxon is shown in the legend. The number of sequences (PLSDB plasmids and UniParc proteins)
1181 associated with each leaf is indicated by lines in the outmost rings. The line length represents
1182 the number of sequences but the upper limit in the visualisation is set to 100.

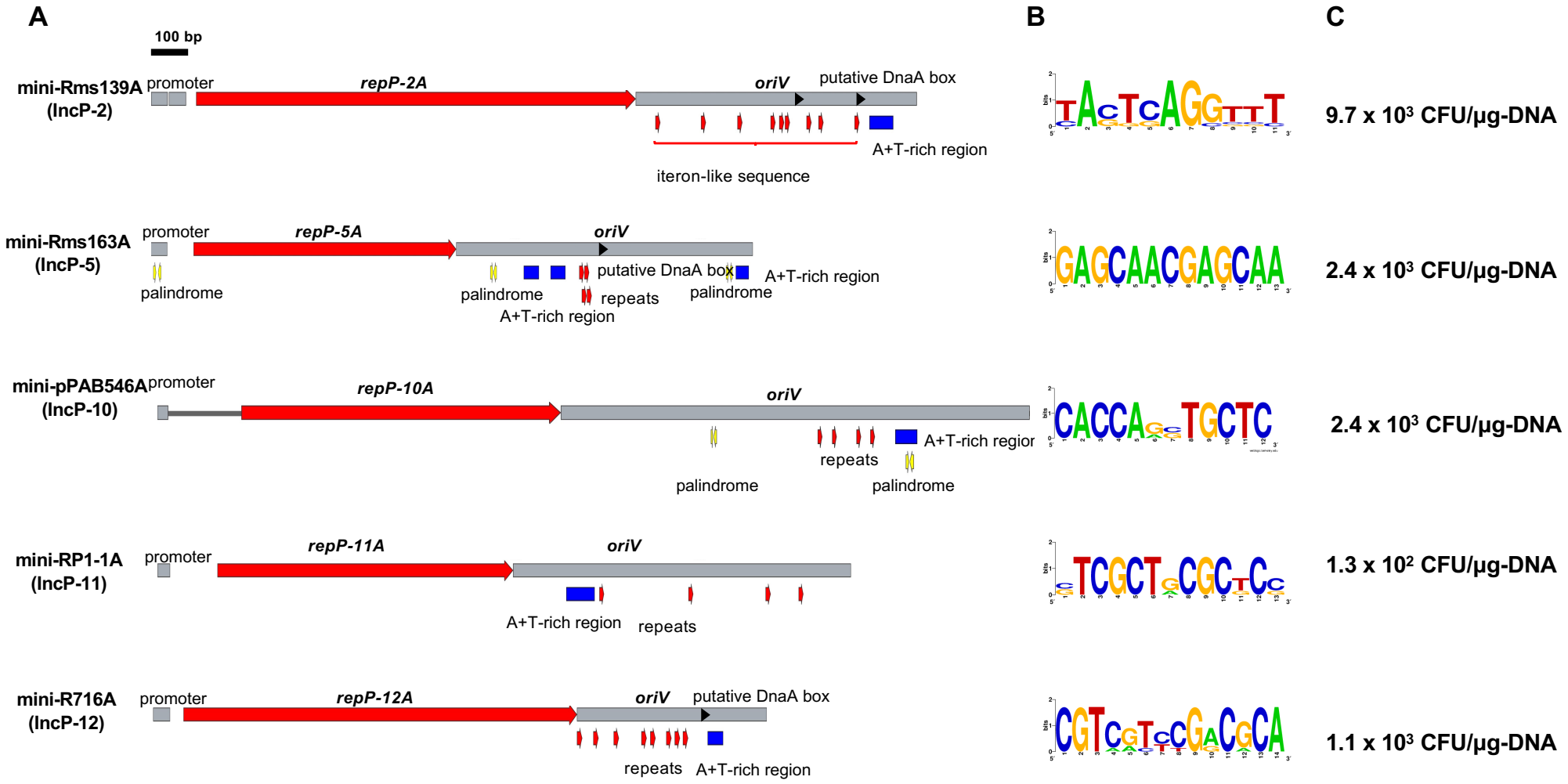
A
B
C
100 bp
mini-Rms139A (IncP-2)
promoter
repP-2A
oriV
putative DnaA box
A+T-rich region
iteron-like sequence
9.7 x 10³ CFU/µg-DNA
mini-Rms163A (IncP-5)
promoter
repP-5A
oriV
palindrome
palindrome
putative DnaA box
A+T-rich region
repeats
palindrome
A+T-rich region
2.4 x 10³ CFU/µg-DNA
mini-pPAB546A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

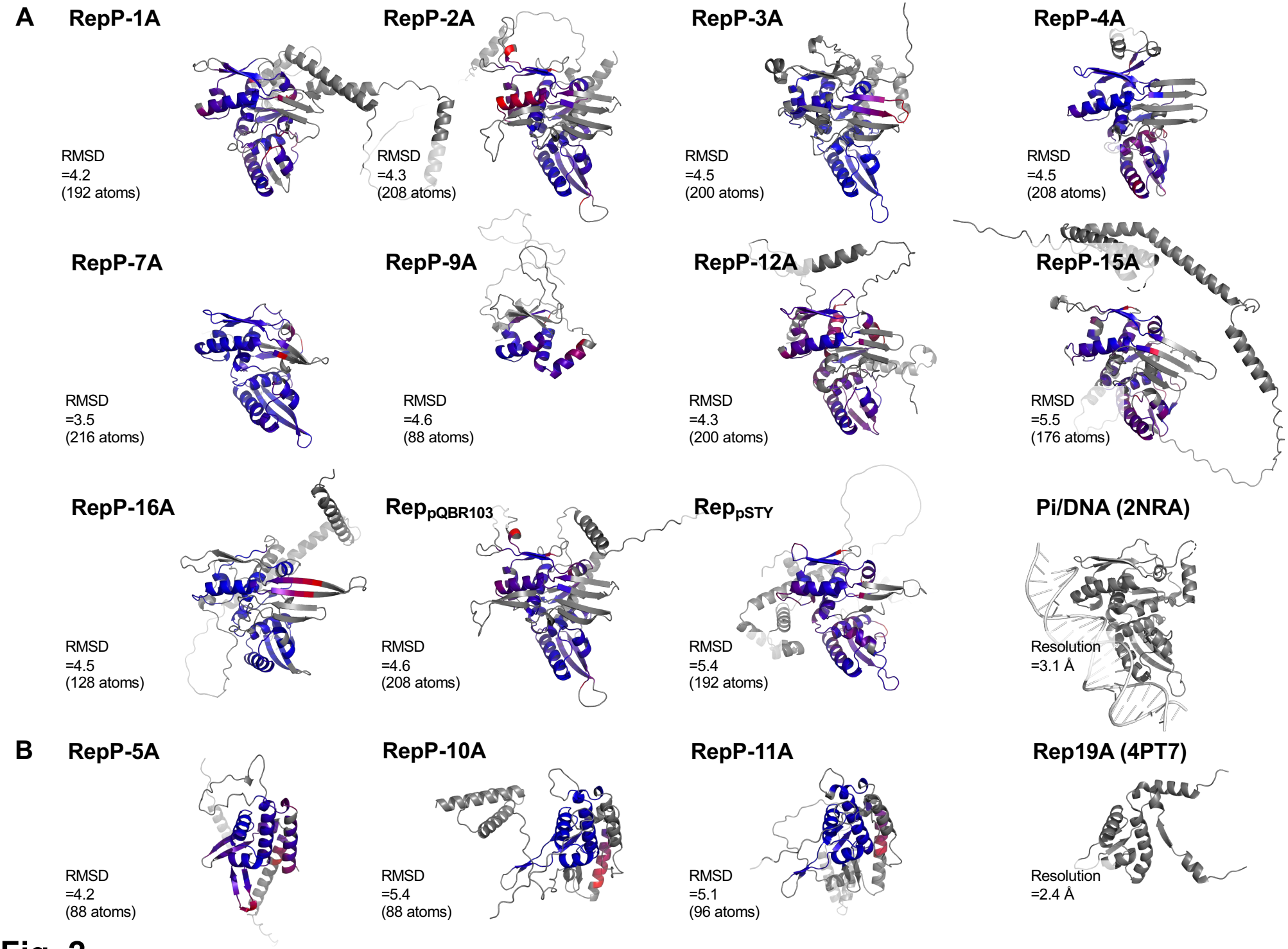
A
RepP-1A
RMSD =4.2 (192 atoms)
RepP-2A
RMSD =4.3 (208 atoms)
RepP-3A
RMSD =4.5 (200 atoms)
RepP-4A
RMSD =4.5 (208 atoms)
RepP-7A
RMSD =3.5 (216 atoms)
RepP-9A
RMSD =4.6 (88 atoms)
RepP-12A
RMSD =4.3 (200 atoms)
RepP-15A
RMSD =5.5 (176 atoms)
RepP-16A
RMSD =4.5 (128 atoms)
Rep pQBR103
RMSD =4.6 (208 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

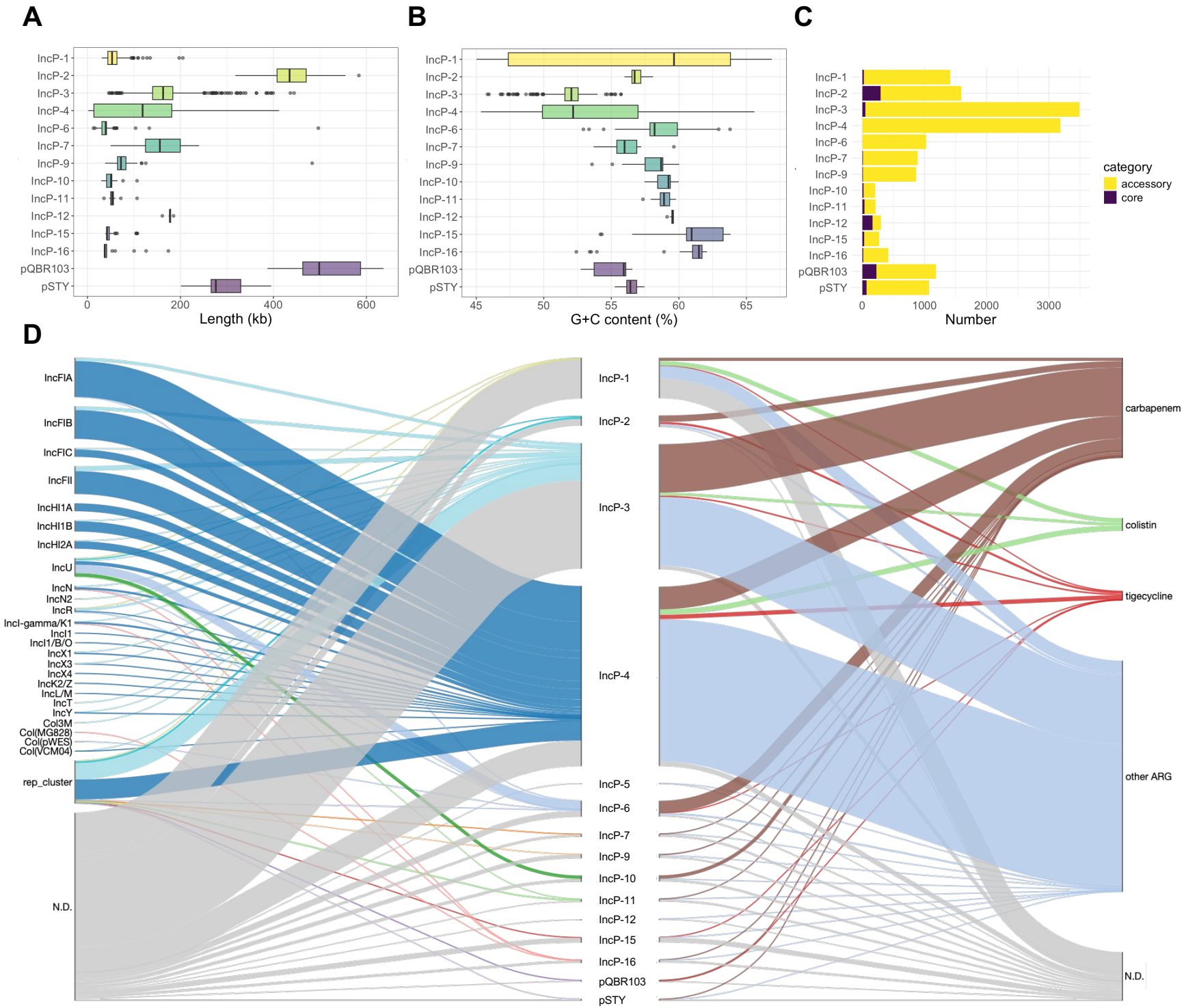
A
IncP-1
IncP-2
IncP-3
IncP-4
IncP-6
IncP-7
IncP-9
IncP-10
IncP-11
IncP-12
IncP-15
IncP-16
pQBR103
pSTY
0
200
400
600
Length (kb)
B
45
50
55
60
65
G+C content (%)
C
0
1000
2000
3000
Number
category
accessory
core
D
IncFIA
IncFIB
IncFIC
IncFII
IncHI1A
IncHI1B
IncHI2A
IncU
IncN
IncN2
IncR
IncI-gamma/K1
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

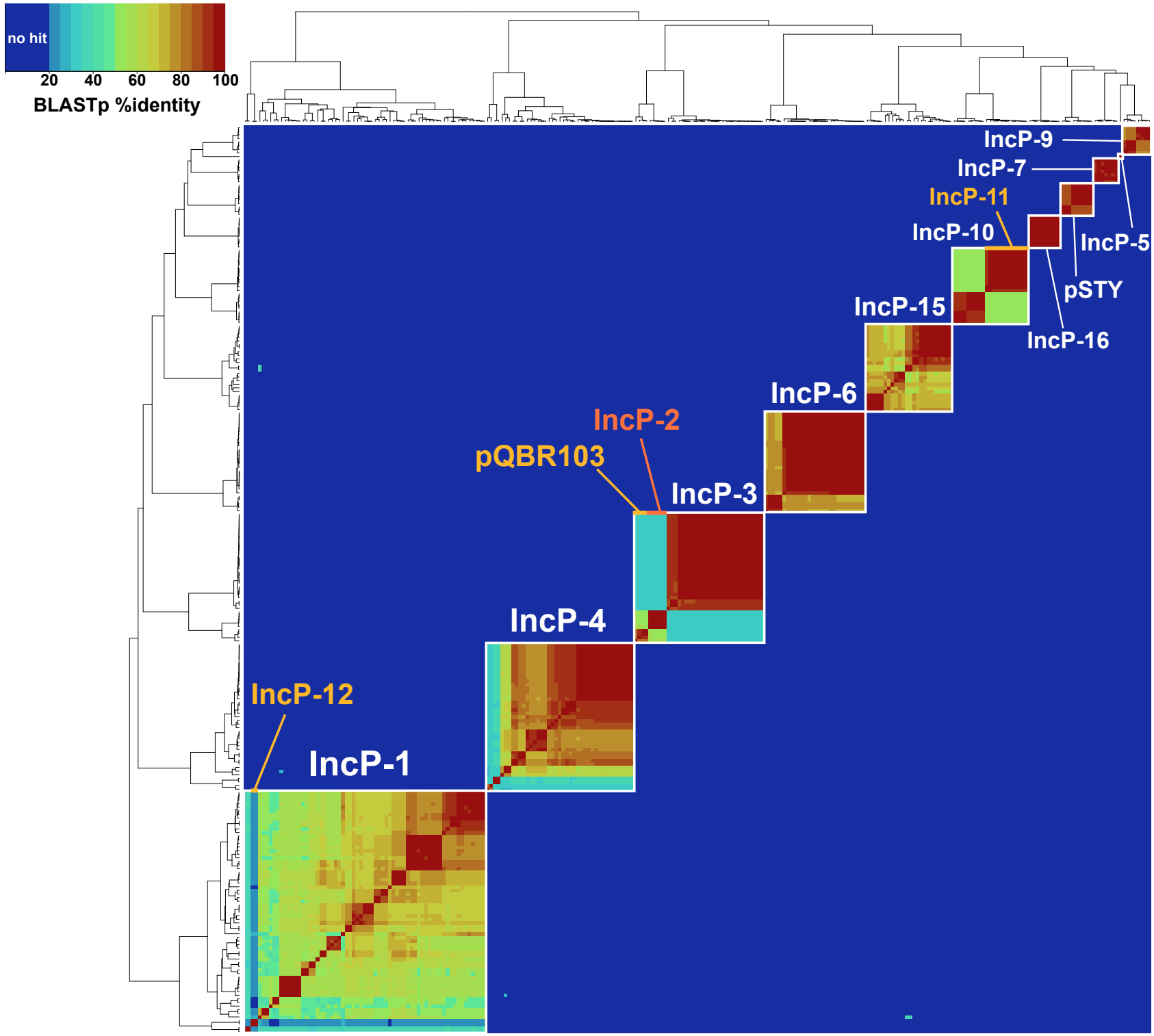
no hit
20
40
60
80
100
BLASTp %identity
IncP-9
IncP-7
IncP-11
IncP-10
IncP-5
pSTY
IncP-15
IncP-16
IncP-6
IncP-2
pQBR103
IncP-3
IncP-4
IncP-12
IncP-1

Fig. 4

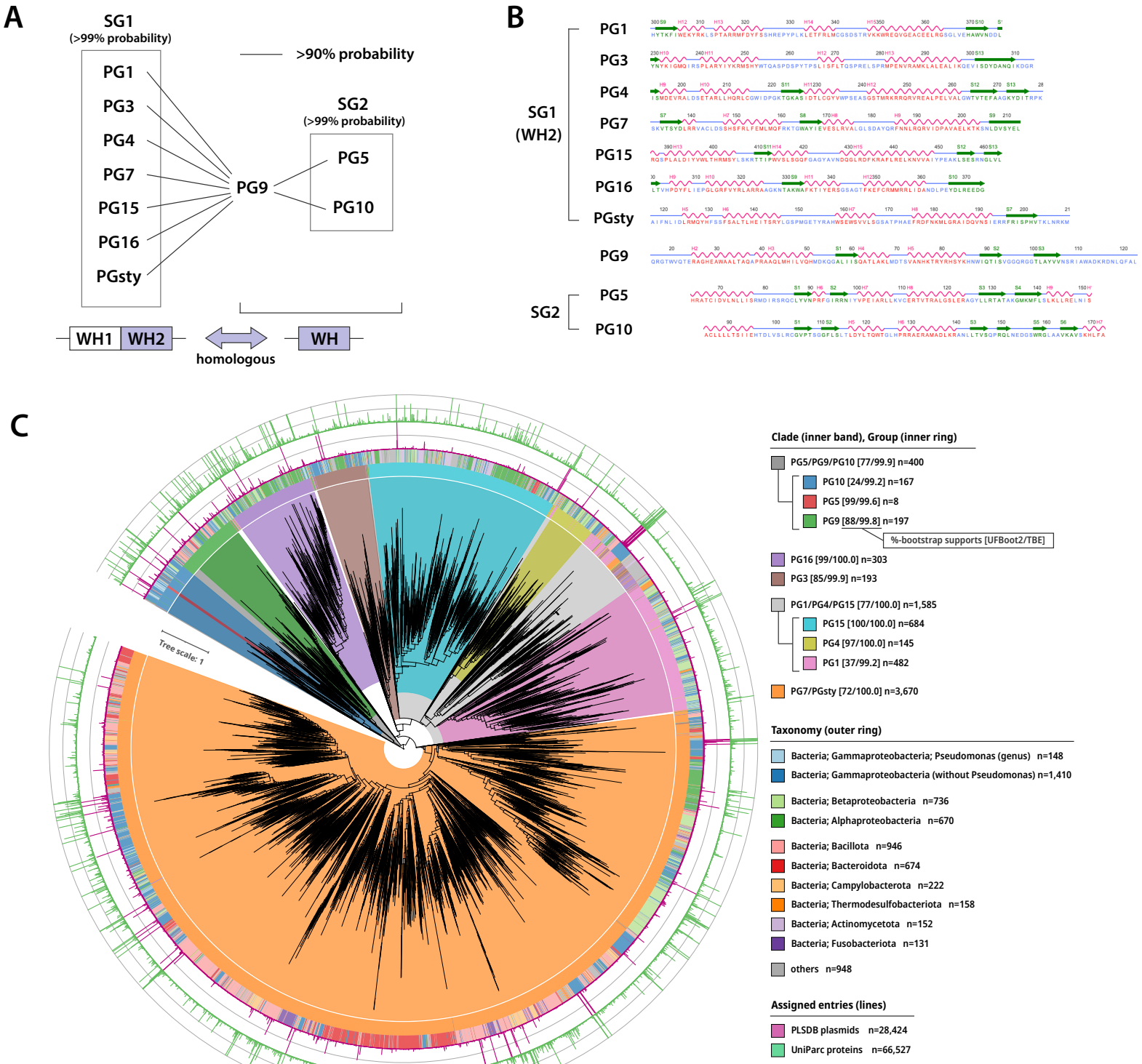
A
SG1
(>99% probability)
PG1
PG3
PG4
PG7
PG15
PG16
PGsty
>90% probability
PG9
SG2
(>99% probability)
PG5
PG10
WH1
WH2
homologous
WH
B
SG1
(WH2)
PG1
PG3
PG4
PG7
PG15
PG16
PGsty
PG9
SG2
PG5
PG10
C
Tree scale: 1
Clade (inner band), Group (inner ring)
PG5/PG9/PG10 [77/99.9] n=400
PG10 [24/99.2] n=167
PG5 [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