A large-scale phylogeny of replication initiation proteins illuminates plasmid macroevolutionary landscape

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

[1] Research Center for Bioscience and Nanoscience (CeBN), Japan Agency for Marine Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka-city, Kanagawa, 237-0061, Japan
[2] Department of Engineering, Graduate School of Integrated Science and Technology, Shizuoka University, 3-5-1 Johoku, Chuo-ku, Hamamatsu, Shizuoka, 432-8561, Japan
[3] Agro-Biotechnology Research Center (AgTECH), Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo, 113-8657, Japan
[4] Department of Clinical Laboratory Medicine, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8507, Japan
[5] Antimicrobial Resistance Research Center, National Institute of Infectious Diseases, Japan Institute for Health Security, 4-2-1 Aobacho, Higashi-murayama, Tokyo 189-0002, Japan
[6] Shizuoka Instrumental Analysis Center, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka, 422-8529, Japan
[7] Research Institute of Green Science and Technology, Shizuoka University, 3-5-1, Johoku, Chuo-ku, Hamamatsu, Shizuoka, 432-8561, Japan
[8] Department of Science, Graduate School of Integrated Science and Technology, Shizuoka University, 3-5-1, Johoku, Chuo-ku, Hamamatsu, Shizuoka, 432-8561, Japan
[9] Faculty of Environment and Information Studies, Keio University, 5322 Endo, Fujisawa, Kanagawa 252-0882, Japan
[10] Japan Collection of Microorganisms, RIKEN BioResource Research Center, 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

† Joint authors

* To whom correspondence should be addressed.

Tel:+81-46-867-9084; Email: nishimuray@jamstec.go.jp (Yosuke Nishimura);
Tel:+81-466-47-5099; Email: haruo@sfc.keio.ac.jp (Haruo Suzuki);
Tel:+81-3-5285-1111; Email: suzuki-m@niid.go.jp (Masato Suzuki);
Tel:+81-53-478-1181; Email: shintani.masaki@shizuoka.ac.jp (Masaki Shintani)

ABSTRACT

Plasmids are key for the acquisition of new traits in bacterial cells, including antimicrobial resistance. The *rep* gene and its encoding replication initiation protein (RIP) are essential for plasmid autonomous replication. Here, we reconstructed a large-scale phylogeny of RIPs containing a conserved winged-helix domain (WH), encompassing 34,724 plasmids including 5,369 yet unclassified ones, together with RIPs likely derived from unrecognized plasmids. The phylogeny incorporates macroevolutionary perspectives into plasmid replication systems and resolves evolutionary relationships among replicon types including existing and newly defined ones. Integration of the knowledge of genetic organization, taxonomic and environmental distribution of plasmids, RIP sequences illuminates the diversification history of plasmids and their replication proteins, including an early event of domain arrangement of WH RIPs. The WH RIP phylogeny opens a new door for plasmid macroclassification centered on the replication system and provides valuable insights into plasmid evolution. (143 words)

Main

Autonomously replicating mobile genetic elements (MGEs), such as plasmids and viruses, drive bacterial evolution through horizontal gene transfer, providing new traits that facilitate bacterial adaptation [1,2]. MGEs, particularly plasmids, have been highlighted as a major vehicle for the spread of antimicrobial resistance genes (ARGs), posing a threat to global public health[1]. Taxonomic classifications of MGEs are essential to gain insight into bacterial evolution and antimicrobial resistance[2,3]. Viral classification has long been developed centred on the International Committee on Taxonomy of Viruses [3], with a hierarchical taxonomic scheme ranging from realms to species [4]. On the other hand, plasmid taxonomy has begun to be proposed [5,6], but still lacks a feasible way and lacks an organised process for discussion and standardisation among researchers in the field.

Historically, plasmid classification has been based on incompatibility (Inc) grouping, which is determined by phenotypic interactions where plasmids of the same group cannot co-exist in a cell mainly due to their similar replication systems[4]. R-factors, later called R plasmids, where 'R' means resistant, have been identified in, for example, *Pseudomonas* (*Pseudomonas aeruginosa*) since the 1970s, with proposed Inc groups from IncP-1 to IncP-14[7]. Although molecular-based knowledge rooted in *Pseudomonas* Inc grouping has been accumulated alongside the Inc groups, sequence-based knowledge remains inadequate, as certain *Pseudomonas* Inc groups still lack plasmid sequences.

Currently, sequence-based plasmid classifications [7], which take over some groups and their names of the Inc grouping, have been established. However, the following limitations remain. (i) Most classifications cannot incorporate macroevolutionary perspectives, due to the dependence on nucleotide sequence similarity, which is not suitable for remote homology detection, such as replicon typing based on the RIP genes and *oriV*s[8] and whole-sequence-based typing, such as PTUs (plasmid taxonomic units)[9,10]. (ii) Classifications based on the relaxase gene [5,11,12] can be applied only to the subset of plasmids that are mobilizable, as relaxase genes were detected in only 42% of plasmids in a recent dataset[7]. Overall, the current situation is still distant from the establishment of a universal macroclassification of plasmids, which is essential to understand the evolutionary history of plasmids from their origins to the current diversified state of plasmids.

Here, to implement a new plasmid classification that incorporates widely-used replicon types and the macroevolutionary landscape, a new replicon sequence-based classification, named PInc(s), was developed by supplementing sequences of *Pseudomonas* plasmids and thoroughly identifying the replicon-related sequences. Most RIPs from PIncs, termed WH RIPs, were found to contain a remotely homologous winged-helix (WH) domain, suggesting a common evolutionary origin. To reconstruct macroevolutionary trajectory of WH RIPs, a large-scale phylogeny incorporates WH RIPs from 34,724 plasmids, together with RIPs likely derived from unknown plasmids, into macroevolutionary perspectives back to the ancestral plasmid evolution suggesting an early event of domain arrangement of the WH RIPs.

Results

Sequencing and molecular characterisation of *Pseudomonas* plasmids

The complete nucleotide sequences of six plasmids were determined, namely Rms139 (IncP-2)[13], Rms163 (IncP-5)[14,15], Rsu2 (IncP-9)[16,17], RP1-1 (IncP-11)[18], R716 (IncP-12)[19], and pMG26 (IncP-13)[20,21] (Table 1). For IncP-10 plasmids, since only a partial plasmid sequence of the IncP-10 archetype plasmid (R91-5[22,23]) is available, we selected pPAB546 as a representative (Supplemental Text S1). A sequencing assembly of pMG26, which was previously classified as IncP-13[24], resulted in no circular contigs after removing reads that were mapped to the *P. aeruginosa* PAO1 chromosome. Based on detailed molecular characterisation, we concluded that pMG26 is not a plasmid but an integrative and conjugative element (ICE), and therefore IncP-13 plasmids are considered ICEs (Supplemental Text S1).

The RIPs and *oriVs* of the plasmids (Table 1) were identified by sequence- and structure-based searches and functionally validated using mini-replicon constructs (see Methods, Figure S1, S2, and Supplemental Text S1). In addition, transferability of five plasmids (Rms139, Rms169, Rsu2, RP1-1, and R716) was experimentally tested by mating assays on solid and liquid conditions (Supplemental Text S1). All plasmids were successfully transferred from *P. aeruginosa* PAO1808 to *P. aeruginosa* PAO1 in filter mating assays, confirming their self-transmissibility (Table S1).

No transconjugants of Rms163 were detected in the liquid mating assay (Table S1). Based on the MOB-suite analysis, T4SS family and a MOB type of Rms139 (IncP-2) and Rsu2 (IncP-9), RP1-1 (IncP-11) could be predicted, whereas Rms163 (IncP-5) and R716 (IncP-12) harbored Dot/Icm secretion systems, which differ from the traditional type IV secretion systems (T4SS)[25]. These different conjugation systems and their features possibly reflect environmental conditions[26], indicating that each plasmid may have a preference for certain environmental conditions that make its transfer more efficient.

PIncs classify unclassified plasmids in PLSDB

A new replicon typing classification named PInc(s) was developed (Table 1). PIncs rooted in the *Pseudomonas* Inc grouping and inherits numbering of the Inc grouping, such as PInc-1 (originated from IncP-1), and/or PInc-2 (from IncP-2). Some Inc groups were excluded from PIncs, due to non-plasmid characteristics for IncP-8[27] and IncP-13 (this study), or material unavailability for IncP-14[28]. Additionally, certain plasmid groups of *Pseudomonas* were classified in PIncs that have been reported to disseminate antimicrobial resistance but are not included in the Inc grouping, namely PromA[29], pSN1216-29-like[30], pQBR103-like[31], and pSTY-like[32] groups, defined as PIncs as PInc-15, PInc-16, PInc-17, and PInc-18, respectively (Table 1, Supplemental Text S1). PIncs subgroups were defined according to subgroups of the Inc groups or newly defined if a group contains a large nucleotide diversity of the RIP genes. Plasmid sequences were collected and the RIP genes in the plasmids were identified based on sequence and structural homology to known RIPs (Figure 1). Reference plasmids (n=87), including the newly sequenced ones, were identified covering the 15 PInc groups and 22 subgroups (Table S7). The reference RIP sequences (n=87), as a reference for the group/subgroup definition, were identified from the reference plasmids (figshare). Each PInc group is defined as follows: a collection of plasmids that has a RIP gene showing nucleotide (BLASTn) similarity to a reference RIP of the group above 90% identity and 60% length coverage, considering only the best hit among the reference RIPs. The subgroups were defined equivalently.

Plasmids in PLSDB (n=59,895) were searched for members of PIncs using the reference RIPs. The search identified 2,357 plasmids classified in the PInc groups (Tables S2), including 530 plasmids previously unclassified by PlasmidFinder and 1,514 by MOB-suite. The plasmids are hosted not only by *Pseudomonas* but also by various strains of *Enterobacterales*, possibly due to the presence of broad host range plasmids (https://figshare.com/s/915a336484ee084ffbea). During the classification process, we identified that 109 plasmids were potentially misclassified in the MOB-suite database, because the IncP-6, IncP-7, and IncP-9 plasmids were classified as 'IncP' of *Enterobacterales* Inc grouping, which is considered equivalent to IncP-1 of the *Pseudomonas* Inc grouping[33]. Overall, PIncs improved accuracy and coverage of the plasmid classification.

Pangenomic characteristics of PIncs

PIncs allows large-scale pangenome analyses to explore relationships between the replicon types and the diversity of plasmid genomes (Tables S2, S3-1~S3-14). Detailed pangenomic characterisation for each PInc was provided in Supplemental Text S1. A comprehensive identification of the RIP gene using several tools revealed that over half of PInc-4 (70.9%), PInc-11 (65.2%), PInc-6 (60.5%), and PInc-10 (51.0%) plasmids are multi-replicon (Figure 2D, Table S2). By contrast, less than 10% of PInc-1, PInc-5, PInc-7 , PInc-9, and PInc-15 plasmids are multi-replicon. These rates of multi-replicons are not attributable to the average genome size of the PInc groups, reflecting that each PInc has a distinct history of the replicon coupling. Specifically, PInc-4 plasmids are coupled to many types of replicons (Figure 2D, Table S2). In contrast, most multi-replicons of PInc-10 and PInc-11 were coupled to each other or to only a few types of replicons (Table S2).

The genome sizes of PInc-2, PInc-17, and PInc-18 plasmids were exceptionally large, while those of PInc-3 and PInc-4 showed wider ranges compared to others (Figure 2A). The wider G+C content range in PInc-1 and PInc-4 plasmids (Figure 2B) may reflect amelioration to diverse bacterial hosts with different G+C content [34].

Core genes present in 95–100% plasmids in each PInc were identified using PIRATE. PInc-2, PInc-12, and PInc-17 plasmids had more core genes than other groups (Figure 2C). In contrast, PInc-1 plasmids, found in various environments[35], likely exhibit higher diversity.

Evolutionary insights into RIPs

Homolog collection To understand the evolutionary relationships of RIPs, protein homology between 251 unique RIP sequences from PIncs plasmids was identified using

BLASTp (Figure S4). Each PInc group shows a different degree of sequence similarity of RIPs, some PIncs showing high sequence variation, namely PInc-1 (34.05% identity for the most divergent pair), PInc-4 (28.81%), and PInc-15 (56.14%). In contrast, the other groups contain less diverse RIP proteins (>73% identity within each PInc) (Table S4). RIPs of different groups show sequence homology for specific combinations such as PInc-2/PInc-3/PInc-17, PInc-1/PInc-12, and PInc-10/PInc-11, suggesting the evolutionary relatedness of these plasmids (Figures S3 and S4).

To explore protein homology beyond the BLASTp sensitivity, diverse RIP homologs for each PInc were collected by an iterative search using RIPs of PIncs and other 8 known RIPs as queries. In addition, other known RIPs were used as queries to complement the coverage of the RIP collection. The detailed procedures are written in Supplemental Text S1 and Figure S4. In short, homologous sequences were iteratively collected from PLSDB and UniParc, identified 5,488 representative sequences from RIP protein clusters at 50% identity (63,179 sequences in total), which were classified into 14 homologous groups.

Remote homology detection Remote homology of the RIPs between the 14 groups was identified with HHblits[36], which performs sequence-based hypersensitive homology search employing an hmm-hmm comparison. With the exception of G-PInc-6 (the group including PInc-6), the RIPs of the remaining 13 groups exhibited homology (Figure S5). Two supergroups, 'sWH', which contain a single winged-helix (WH) domain, including 7 groups and 'dWH', which contain double WH domains in tandem, including 6 groups, were identified, according to homology within each supergroup being significant (>95% probability for 30 out of 36 all possible combinations of sWH groups; >95% probability for all possible combinations of dWH groups) (Figure S5). Notably, homology between sWH and dWH groups were identified, 15 out of 42 (i.e., 6 x 7) possible combinations of groups being significant (>80% probability) and 21 combinations being significant if bit less significant hits are included (>50% probability).

The solved and predicted crystal structures of dWH RIPs, such as RepA from plasmid pPS10 (PG7), RepE from plasmid F (PG7), Pi from plasmid R6K (PG7), and TrfA protein from plasmid RK2 (PG1), contain two winged-helix domains (WH1 and WH2)[37], and the two WH domains are included in the identified conserved regions among dWH. Although only crystal structures of two RIPs (pSK41 and pTZ2162 plasmids) of sWH have been reported[38], Alphafold3[39] predictions suggest that they have only one WH domain (Figure 1). The homologous region for groups of sWH and dWH contains the WH domain of sWH groups and the WH2 domain (the C-term WH domain) of dWH groups (Figure S5). It was reported that the WH2 domain of the RepA protein of pPS10 plasmid binds precisely to the 3' end of the iteron and is the major determinant of

RepA sequence specificity[45]. Together with their broad conservation, the conserved WH domain likely plays an important role in replication origin recognition of plasmids. Here, we named the conserved WH domain-containing RIPs 'WH RIPs', which constitute a 'WH RIP superfamily'. On the other hand, it has been proposed that the WH1 domain (the N-term WH domain) is evolutionarily related with the WH domain of the archaeal Orc1/Cdc6 gene and the eukaryotic Orc genes, both of which encodes origin recognition proteins for the chromosomal replication[40].

RIPs of G-PInc-6 showed no homology with the other groups (<1% probability by HHblits). To characterise the evolutionary origin, an HHblits search was performed against Pfam v35.0. RIPs of G-PInc-6 show significant homology to PF03090 (replicase family; with 98.4% probability), including the ColE2 Rep protein, which is known to be the only plasmid-specific trans-acting factor for initiation of plasmid replication[41] and belongs to the superfamily of archaeo-eukaryotic primases (AEPs)[42]. The RIPs of PInc-6 plasmids[43,44] have an additional PriCT domain that plays a central role in DNA unwinding. We therefore concluded that the G-PInc-6 RIPs have different evolutionary origins from WH RIPs, and named 'AEP RIPs'.

Phylogenetic reconstruction A phylogeny of WH RIPs from the 13 groups were reconstructed using the identified homologous region (Figure 3A). The maximum likelihood phylogenetic tree consists of 4,769 representative sequences of protein clusters at 50% identity, associated with 34,724 PLSDB plasmids and 95,057 UniParc proteins (Figure S4B). We defined 8 clades and 42 subclades based on the branch support values (Figure 3B). All the clades and subclades were supported by either or both of the high support values (UFBoot2: ≥95%, SH-aLRT: ≥80%). As an overall tree structure, sWH RIPs and dWH RIPs have only one split within the tree, which is highly supported (UFBoot2: 89.4%, SH-aLRT: 88%). This split is highly likely to correlate with an ancestral event that causes the acquisition of an additional WH domain or conversely, the loss of the domain, which is a highly crucial conversion for the molecular evolution of WH RIPs.

The phylogenetic tree associates many sequences from PLSDB and UniParc. A large part (72.7%; n=3,466) of the leaves are associated only with UniParc sequences and no association with RIPs of PLSDB plasmids, suggesting the existence of numerous unknown plasmids. In contrast, almost half (57.2%; n=34,724) of PLSDB plasmids are associated with the remaining 1,303 tree leaves. Considering plasmids using alternative replication mechanisms, such as the superfamily of AEPs (e.g., G-PIncP-6) and rolling-circle replication (e.g., Pfam PF01719: Rep_OBD), the fraction included in the WH phylogeny (57.2%) is notably high. Notably, the tree associates with RIPs of 5,369 PLSDB plasmids that do not contain RIPs identified with PlasmidFinder, MOB-suite, and

AcinetobacterPlasmidTyping[45]. Therefore, these 5,369 plasmids were newly characterized based on their RIP sequences. Among the 1,303 leaves associated with PLSDB plasmids, 59 leaves are associated with more than 100 plasmids, collectively accounting for RIPs of 27,654 plasmids (79.6% of the 34,724 PLSDB plasmids associated with the tree). This highlights that PLSDB plasmids are enriched with ones having similar RIP sequences. Moreover, 11,570 PLSDB plasmids (33.3% of the 34,724 PLSDB plasmids associated with the tree) identified as multi-replicons that contain multiple WH RIPs, suggesting that plasmids containing multiple WH RIPs are common.

Known replicon typings, including PIncs, are mapped onto the phylogenetic tree and the positions are widely distributed (Figure 3B). RIPs from 2,205 PIncs plasmids are associated with 15 leaves that are distributed in 7 clades (except the clade D). The RIPs of PInc-1, PInc-4, and PInc-15, which compose relatively high sequence diversity, are associated with multiple leaves. On the other hand, RIPs from the other PIncs are associated with a single leaf, reflecting the high degree of similarity (>73% of protein sequence identity; Figure S3). Moreover, some combinations of multiple PIncs were merged into one leaf (i.e., PInc-2/-17, PInc-10/-11), reflecting the higher sequence similarity beyond the PIncs.

Each clade has characteristic taxonomic and environmental distribution. Most clades, especially clades of dWH groups, contain diverse taxonomic lineages, suggesting potential host transitions within the clade (Figure 3A). In contrast, some clades show biased taxonomic distribution, such as *Bacillota* being predominant in the clades B and D and *Gammaproteobacteria* in the clade C. More than half of the RIPs of *Alphaproteobacteria*, *Planktomycetota*, and *Deinococcota* plasmids belong to the clade E. Most RIPs of *Actinobacteria* and *Acidobacteria* belong to the clade G (>60%). Most RIPs of *Bacteroidota*, *Campylobacterota*, and *Fusobacteriota* belong to the clade H (>80%). These taxonomic distributions highly likely reflect the evolutionary origin and trajectory of the host lineage transition of the plasmid.

For environmental distribution, sWH groups (the clades A-D) show a high proportion of host-associated microbiomes. The clades B and D show enrichment in host-associated microbiomes, especially Human- and Mammals-associated microbiomes. The clades A and C are also enriched with host-associated microbiomes, especially insect-associated microbiomes for A and plant-associated microbiomes for C. Host specificities may also be connected to specific environments, possibly reflecting their ecological or pathogenic roles. On the other hand, sWH groups (the clades A-D) show a high proportion of environmental microbiomes such as aquatic and terrestrial environments. The clades E and F show enrichment in aquatic environments, while the clade G show enrichment in terrestrial environments. Further research is needed to

investigate how these trends of environmental niches are caused during the plasmid and bacterial evolution.

Discussion

The reconstructed large-scale phylogenetic tree of WH RIPs not only provides a valuable resource for future studies of plasmid replication mechanisms and these evolution, but also greatly enhances our understanding of plasmid diversity, including the identification of many RIPs that are likely derived from unknown plasmids. Experimental validation of such candidate RIPs is an essential step to understand yet unidentified plasmids. The newly classified RIPs containing plasmids based on the phylogenetic tree, including unknown plasmids, will provide a clue for characterising the vector of gene dissemination such as ARGs, among bacterial populations.

The WH RIPs phylogeny provides new insights into plasmid evolution. First, the tree structure strongly suggests an ancient event that transformed the domain structure of RIPs between the form of double WH domains (dWH) and single WH domain (sWH). The evolutionary origin of the WH2 domain has rarely been discussed, in contrast to the WH1 domain[46] and we first identified the homologous relationship between the WH2 domain and the WH domain of sWH. Second, the taxonomic enrichment of several lineages in the tree would provide a clue to understand the complicated path of host transition history through the plasmid evolution. Third, the phylogeny illuminates relationships of WH RIPs including ones from PIncs, providing a clue to understanding the evolutionary path of the transition of molecular mechanisms for plasmid replication.

The RIP gene is essential for plasmids, as RIP is the only plasmid-encoded protein required for autonomous replication of plasmids. Although the RIP gene has been used as a genetic marker for replicon typing, the development of plasmid macroclassification, or 'plasmid taxonomy', based on a RIP phylogeny has been criticised for two issues[5]. (i) Some plasmids have multiple loci for autonomous replication and thus contain multiple RIPs. (ii) RIPs would belong to a large number of protein families, reflecting many types of mechanisms for plasmid replication. Here, we propose a feasible path for RIP phylogeny-based plasmid taxonomy development. For the first issue, multiple assignments are applicable by assigning a multi-replicon plasmid to multiple taxa based on each RIP sequence. The resulting taxonomy will illuminate the replication system-centred evolutionary history of plasmids, including the 'birth and death' of multi-replicons, from a macroevolutionary perspective.

For the second issue, RIPs surely belong to multiple protein families. There are multiple replication mechanisms for circular plasmids, named as theta type, strand

displacement, and rolling circle[47]. RIPs from PIncs plasmids are derived from two evolutionarily distinct protein families, namely WH RIPs and AEPs, and likely function in the theta type or strand displacement replication. In addition, some RIPs for the rolling circle replication are known to belong to the HUH endonuclease superfamily[48] and others show similarity to Rep_trans (PF02486 in Pfam) which belong to another protein family. Since the current viral taxonomy encompasses evolutionarily independent six realms[49], the analogous scheme is feasible for plasmid taxonomy, where the highest rank is divided based on evolutionary independent RIP superfamilies. The lower ranks in the taxonomic hierarchy could be subsequently defined based on phylogenetic trees of the RIP superfamilies, with analogous approaches being taken for viruses[50] and prokaryotes[51].

Overall, the phylogenetic trees covering all families of RIPs will serve as the basis for establishing universal plasmid macrotaxonomy that encompasses the evolutionary history of plasmid replication systems, therefore enhancing our ability to track bacterial evolution and manage antimicrobial resistance in the environment, ultimately contributing to better public health outcomes.

Methods

Bacterial strains, plasmids, and culture conditions

The bacterial strains and plasmids used are listed in Table S5. *E. coli* and *Pseudomonas* strains were cultured in Luria broth (LB) or M9 broth[52] at 30°C (*P. putida* KT2440) or 37°C (*P. aeruginosa* PAO1 and its derivatives). LB plates containing 1.5% agar were used for filter mating. Ampicillin (Ap, 50 µg/mL), piperacillin (Pip, 16 µg/mL), rifampicin (Rif, 50 µg/mL), streptomycin (Sm 150 µg/mL), and tetracycline (Tc, 50 µg/mL for the others) were added to the medium.

DNA manipulation

Total DNA from the bacterial strains was extracted using a NucleoSpin Tissue Kit (TAKARA BIO, Japan). Polymerase chain reaction (PCR) was performed on a T100 thermal cycler (Bio-Rad Laboratories, USA) using the KOD One PCR Master Mix (TOYOBO, Japan). The primers used in this study are listed in Table S6. The following amplification conditions were used: 30 cycles at 98°C for 10 s, 67°C for 5 s, and 68°C for 1 s. The pUC-GW plasmid with DNA fragments including putative RIP genes and *oriV*s of each plasmid, except those for RP1-1 and pMG26, was synthesised by the custom service of Azenta Life Science, USA. Restriction enzymes (New England Biolabs, USA or TAKARA BIO), the HiYieldTM Gel/PCR DNA fragments Extraction kit (RBC Bioscience, Taiwan), NEBuilder Hifi DNA Assembly system (New England Biolabs), and competent *E.*

coli JM109 and DH5α cells (RBC Bioscience) were used for cloning the DNA fragments. Plasmid DNA from the bacterial strains was extracted using the NucleoSpin Plasmid EasyPure Kit (TAKARA BIO). Electroporation to introduce the plasmids into different bacterial strains was performed using a MicroPulser electroporator (Bio-Rad Laboratories). All other procedures were performed according to standard methods[52].

Conjugation assays

Filter mating assays were performed as described previously[53]. In brief, donor and recipient strains, *P. aeruginosa* PAO1808 with a plasmid and *P. aeruginosa* PAO1R or *P. putida* KT2440, were cultured in LB with the appropriate antibiotic(s) (preculture). One millilitre of the above culture in LB medium was harvested and suspended in fresh LB medium. Approximately 10^8 colony forming units/mL (CFU/mL) of the donor and recipient suspended in 150 µL LB were mixed and then dropped on 0.2 µm pore-size filters (cellulose acetate, Advantec) on an LB agar plate for 24 h at 37°C. The mixture on the filter was resuspended in PBS and then spread onto LB plates with appropriate antimicrobials to select transconjugants (i.e., recipient cells with the plasmid). Liquid mating assays were performed as follows. The precultured donor cells were resuspended in 500 µL LB, mixed with 1 mL recipient cells for 2 h (200 rpm), and statically incubated for 3 h at 37°C. The cells were then harvested, resuspended in PBS, and spread onto LB plates with the appropriate antibiotics to select the transconjugants.

Complete sequencing of the plasmids by next-generation sequencing

The nucleotide sequences of the plasmid DNAs were determined as follows: whole genomic DNA extracted from their host (*P. aeruginosa* PAO1808, a derivative strain of PAO1[54]) was sequenced using the HiSeq2500 platform (Illumina, USA). Trimmed high-quality short reads (read length >140 bp and quality score >15) were assembled using SPAdes v3.14.0 using the plasmid option[55]. If circular contig(s) could not be found, the host chromosomal DNA was removed by mapping the resultant contigs on the host genome sequences [their deposited sequences, that is, DDBJ/GenBank accession number NC_002516 (*P. aeruginosa* PAO1)] using Geneious Prime software v2020.01[56]. Subsequently, the plasmid reads were extracted from the remaining contigs using SAMtools v1.7[57] and SeqKit v0.8.0[58] and reassembled using SPAdes[59]. For RP1-1 and pMG26, gaps in the plasmids were closed by PCR with the designed primer sets based on the ends of contigs and Sanger sequencing of the PCR products.

Gene annotation of plasmids

Gene annotations of the plasmids were performed using DFAST-core v1.2.5[60] with an in-house database of plasmid sequences collected from NCBI RefSeq and then corrected manually. Synteny analyses of plasmids were performed using Easyfig v2.2.2[61]. Other *in silico* analyses were performed using Geneious Prime v2020.01. The annotated plasmid sequences were deposited under the following GenBank accession numbers: LC685025, LC685026, LC685027, LC685593, and LC700336.

Detection and phenotype identification of ARGs

ARG sequences were detected using Staramr v0.10.0 and ResFinder database v2.2.1 updated on 2023-10-27 (https://bitbucket.org/genomicepidemiology/resfinder_db)[59,60] with the following criteria: ≥ 90% identity and ≥ 60% coverage. Clinically important ARGs, including genes conferring resistance to carbapenem, colistin, or tigecycline, in Gram-negative bacteria were further detected based on the phenotype information of ARGs in AMRFinderPlus v3.11.26 updated on 2023-11-16 (https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMRFinder/)[61].

Prediction of RIP genes and *oriV*s of each plasmid

To identify the genes encoding the RIP and origin of vegetative replication (*oriV*) regions of Rms139, Rms163, RP1-1, and R716, two prediction approaches were used; after coding sequence (CDS) annotations using DFAST-core, protein.faa files generated by DFAST-core were used for the following: (i) nucleotide sequence-based searches were performed using BLASTn; (ii) amino acid sequence-based protein structure prediction and the structure-based searches were performed as described later. The GC-skew value, which has been commonly used to measure the asymmetry in the composition of two nucleotides, guanine (G) and cytosine (C), along a DNA strand and is defined as (G-C)/(G+C), of the plasmids was used by Webskew (https://genskew.csb.univie.ac.at/webskew) to detect the changes in the GC skew, which indicates the DNA replication origins or termini[62,63]. Subsequently, the nucleotide or amino acid sequences of putative CDSs around the above GC-skew shift point(s) were subjected to methods (i) or (ii). The putative promoter region of the RIP gene was predicted using BPROM (http://www.softberry.com/berry.phtml?topic=bprom&group=help&subgroup=gfindb)[64]. Subsequently, the predicted RIP and *oriV* sequences of each PIncs were cloned into the pUC-GW-Amp vector (Azenta Life Science). Transformation of *P. aeruginosa* PAO1808 was performed with each of the resultant plasmids (at most 472 ng/µL), and the cultures were spread onto LB+Pip plates. After incubation at 37ºC overnight, colonies on the plates were isolated. Next, the presence of the mini-plasmid was confirmed by PCR with specific primers after extraction of total DNA of transformants.

Protein structure prediction of RIPs

AlphaFold3-mediated protein structure prediction was performed using the amino acid sequences of the RIP gene products in representative *Pseudomonas* plasmids (https://figshare.com/s/9554d4fc982d7929fb77) on the AlphaFold server[39]. A structure-based protein homology search using the predicted structural models was performed for the PDB100 database on the FoldSeek server in 3Di/AA mode[65]. The predicted and experimental structures were visualised and compared using PyMOL v3.1.3.1. The predicted structures were coloured based on the root mean square deviation (RMSD) from the reference experimental structures, i.e., the Pi or Rep19A[38,66] proteins, which are RIPs of the R6K and pCA-347 plasmids, respectively, using the ColorByRMSD script (https://pymolwiki.org/index.php/ColorByRMSD). The distances between aligned C-alpha atom pairs were stored as B-factors of these residues, which are coloured using a colour spectrum, with blue specifying the minimum pairwise RMSD and red indicating the maximum. Unaligned residues were coloured grey.

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

Complete plasmid sequences were downloaded in FASTA format using the Entrez Programming Utilities (E-utilities; https://www.ncbi.nlm.nih.gov/books/NBK25501/) from a public database (PLSDB 2023_11_03_v2) (https://figshare.com/s/915a336484ee084ffbea)[67,68]. To classify the plasmids in the PLSDB based on their rep sequence similarity, the RIP gene sequences were detected using Staramr v0.10.0 (https://github.com/phac-nml/staramr) and rep databases of MOB-suite v3.1.8 updated on 2023-12-12 (https://github.com/phac-nml/mob-suite) (https://figshare.com/s/233faa319102390218e5)[69], PlasmidFinder v2.1 updated on 2023-12-04 (https://bitbucket.org/genomicepidemiology/plasmidfinder_db) (https://figshare.com/s/7de82d807b4a1630f047)[70], and our custom-made *Pseudomonas* plasmid RIP gene database (termed repP database, https://figshare.com/s/28e3204f6991d6ce0c11) with the following criteria: ≥90% identity and ≥60% coverage. The Tus protein, known for binding to DNA replication terminus (ter) sites of replicons, encoded by rep_cluster_1115 in the MOB-suite database, is not included in our custom-made RIP reference sequences. The repP database was constructed using experimentally identified RIP genes from the 87 reference *Pseudomonas* plasmids shown in Table S7. Furthermore, duplicated RIP sequences between the MOB-suite and repP databases (Table S8) and incorrect RIP sequences, including uncertain IncP and rep_cluster_1115, were manually removed from the MOB-suite database. The modified MOB-suite database (https://figshare.com/s/3e3ced8633136259f678) was used to detect multi-replicon *Pseudomonas* plasmids, which are single plasmids that can belong to more than one

group other than the PIncs groups. Among the candidate IncP-1 plasmid sequences based on the similarity to the RIP sequences, artificially constructed sequences not explicitly labelled as such ("other entries;other sequences;artificial sequences") in the NCBI taxonomy database (https://www.ncbi.nlm.nih.gov/taxonomy) were also present. We identified six IncP-1 plasmids (PD584_L3, PD689_delTT1, pJM105, pJM50, pRK2013, and pUZ8002) through a literature survey, and manually removed these sequences from the putative IncP-1 plasmid sequences.

For the final dataset, the sequence length (bp) and GC content (%) were calculated using SeqKit v2.3.0 (https://bioinf.shenwei.me/seqkit/) [58]. Plasmid sequences belonging to each PIncs group were re-annotated using DFAST-core, with a reference genome from each group to ensure consistency in the functional annotations. GFF3 annotation files produced by DFAST-core were subsequently used as input files for PIRATE, a toolbox for pangenome analysis[71].

Pangenome analysis of plasmids

A pangenome, i.e., the entire set of protein genes from all members of a defined group (all the plasmids of the PIncs group), was analysed using PIRATE (https://github.com/SionBayliss/PIRATE). Here, we used DIAMOND v0.9.14[72] rather than BLAST[73] to identify homologous genes (protein families) in PIRATE. PIRATE was run on default parameters with a high-scoring segment pair (HSP) query length threshold of 0.5, a sequence similarity step range of 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, and 98%, and an inflation value of 2 in MCL (v14.137), a Markov cluster algorithm (https://micans.org/mcl/). The core genes in each PIncs group were those that were present in more than 95% of the plasmids included in the pangenome analysis. Statistical computing and result visualisation were performed using R v4.1.3 (https://www.r-project.org/). The output of SeqKit and PIRATE are available in Figshare (https://figshare.com/s/0f2aa75d1eb4c5848a32).

Homolog collection and phylogenetic analysis of RIPs

Detailed methods are provided in Supplemental Text S1 and illustrated in Figure S6. In brief, to analyse the evolutionary relationships of RIPs, we identified 2,374 RIPs that originated from 2,332 PInc plasmids. Subsequently, 251 non-redundant proteins at 100% identity were extracted using MMseqs cluster v15-6f452[74]. Sequence homology between the non-redundant proteins was assessed using BLASTp[73], and the resultant hits were visualised via CPAP (https://github.com/yosuken/CPAP). To detect remote homology between PIncs and collect comprehensive RIP sequences, we extended the sequence diversity through a multistep homology search process. First, we searched ORFs of PLSDB plasmids predicted using orfM (n = 20,331,985)[75] for homologous RIPs

using hmmsearch of HMMER v3.3.2[76]. In addition, 8 additional RIPs sequences that are distant from RIPs were manually picked up and homology searches using diamond blastp against the ORFs of PLSDB plasmids were performed. This process generated 19 homologous groups of RIPs. Second, UniParc proteins (downloaded on Feb 2024; n = 607,912,929) were searched with hmmsearch using 19 HMMs generated from the homologous groups. We merged the UniParc sequences into 19 homologous groups, and filtered out potential contamination of other functional proteins, including various transcriptional factors, using hmmsearch against the Pfam database. After merging the 19 homologous groups into 14 based on the UniParc sequence overlapping, we selected 5,488 representative sequences for the 14 homologous groups. To investigate whether these groups exhibit homology with each other, a sensitive hidden Markov model HMM–HMM comparison was performed using HHblits v3.3.0[77]. We extracted the homologous regions of the alignments of the 13 groups (containing 6,793 sequences in total). These regions were aligned using MAFFT v7.455[78]. After removing the gappy sequences in the alignment, a maximum-likelihood phylogeny of WH RIPs was reconstructed using the alignment and IQ-TREE v2.2.2.6[79]. We used the parameters "--perturb 0.2 --nstop 500", following the recommendation of the reference manual of the software, in case that many short sequences are included. The best fit model (Q.pfam+F+R10) was selected using ModelFinder[80]. We evaluated the supporting values of the best tree with UFBoot2[81] and SH-aLRT[82] via IQ-TREE. We discarded the abnormally long branches using treeshrink[83]. Overall, the tree structure of sWH RIPs and dWH RIPS exhibits a single, highly supported split (UFBoot2: 89.4%, SH-aLRT: 88%). Since our assumption is that this split represents the root of the tree and corresponds to an ancestral event that led to either the acquisition or the loss of a WH domain, the resultant tree was provisionally rooted by the sWH-dWH split using gotree v0.4.0[84]. The tree-associated PLSDB plasmids, UniParc proteins, and IMG/PR[85] plasmids were identified based on sequence homology search results. Taxonomic affiliation was assigned based on PLSDB plasmids and UniParc proteins, and environmental affiliation was based on IMG/PR plasmids. A detailed information related to the phylogenetic tree is available in Figshare (https://doi.org/10.6084/m9.figshare.28837559).

Data Availability

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

Supplementary Data statement

Supplementary Data are available at online.

Contributions

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 interests

The authors declare no conflict of interest.

References

1. Wiedenbeck, J., and Cohan, F.M. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol. Rev. *35*, 957–976.

2. Haudiquet, M., de Sousa, J.M., Touchon, M., and Rocha, E.P.C. (2022). Selfish, promiscuous and sometimes useful: how mobile genetic elements drive horizontal gene transfer in microbial populations. Philos. Trans. R. Soc. Lond. B Biol. Sci. *377*, 20210234.

3. Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G., and Smith, D.B. (2018). Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res. *46*, D708–D717.

4. Simmonds, P., Adriaenssens, E.M., Zerbini, F.M., Abrescia, N.G.A., Aiewsakun, P., Alfenas-Zerbini, P., Bao, Y., Barylski, J., Drosten, C., Duffy, S., et al. (2023). Four principles to establish a universal virus taxonomy. PLoS Biol. *21*, e3001922.

5. Fernandez-Lopez, R., Redondo, S., Garcillan-Barcia, M.P., and de la Cruz, F. (2017). Towards a taxonomy of conjugative plasmids. Curr. Opin. Microbiol. *38*, 106–113.

6. Redondo-Salvo, S., Fernández-López, R., Ruiz, R., Vielva, L., de Toro, M., Rocha, E.P.C., Garcillán-Barcia, M.P., and de la Cruz, F. (2020). Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. Nat. Commun. *11*, 3602.

7. Garcillán-Barcia, M.P., Redondo-Salvo, S., and de la Cruz, F. (2023). Plasmid classifications. Plasmid *126*, 102684.

8. Carattoli, A., Zankari, E., Garcia-Fernandez, A., Voldby Larsen, M., Lund, O., Villa, L., Moller Aarestrup, F., and Hasman, H. (2014). In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob. Agents Chemother. *58*, 3895–3903.

9. Robertson, J., Bessonov, K., Schonfeld, J., and Nash, J.H.E. (2020). Universal whole-sequence-based plasmid typing and its utility to prediction of host range and epidemiological surveillance. Microb. Genom. *6*. https://doi.org/10.1099/mgen.0.000435.

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

11. Garcillán-Barcia, M.P., Redondo-Salvo, S., Vielva, L., and de la Cruz, F. (2020). MOBscan: Automated annotation of MOB relaxases. Methods Mol. Biol. *2075*,

295–308.

12. Coluzzi, C., Garcillán-Barcia, M.P., de la Cruz, F., and Rocha, E.P.C. (2022). Evolution of Plasmid mobility: Origin and fate of conjugative and nonconjugative plasmids. Mol. Biol. Evol. *39*. https://doi.org/10.1093/molbev/msac115.

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

14. Sagai, H., Hasuda, K., Iyobe, S., Bryan, L.E., Holloway, B.W., and Mitsuhashi, S. (1976). Classification of R Plasmids by Incompatibility in Pseudomonas aeruginosa. Antimicrob. Agents Chemother. *10*, 573–578.

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

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

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

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

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

20. Jacoby, G.A. (1980). PLASMID DETERMINED RESISTANCE TO CARBENICILLIN AND GENTAMICIN IN Pseudomonas aeruginosa. Preprint, https://doi.org/10.1016/b978-0-12-675550-3.50011-3 https://doi.org/10.1016/b978-0-12-675550-3.50011-3.

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

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

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

24. Jacoby, G.A. (1986). Resistance Plasmids of Pseudomonas. Preprint, https://doi.org/10.1016/b978-0-12-307210-8.50013-0 https://doi.org/10.1016/b978-0-12-307210-8.50013-0.

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

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

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

28. Boronin, A.M. (1992). Diversity of Pseudomonas plasmids: To what extent? FEMS Microbiol. Lett. *100*, 461–467.

29. Van der Auwera, G.A., Krol, J.E., Suzuki, H., Foster, B., Van Houdt, R., Brown, C.J., Mergeay, M., and Top, E.M. (2009). Plasmids captured in C. metallidurans CH34: defining the PromA family of broad-host-range plasmids. Antonie Van Leeuwenhoek *96*, 193–204.

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

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

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

33. Thomas, C.M., and Haines, A.S. (2004). Plasmids of the genus Pseudomonas. In Pseudomonas (Springer US), pp. 197–231.

34. Lawrence, J.G., and Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. *44*, 383–397.

35. Hayakawa, M., Tokuda, M., Kaneko, K., Nakamichi, K., Yamamoto, Y., Kamijo, T., Umeki, H., Chiba, R., Yamada, R., Mori, M., et al. (2022). Hitherto-Unnoticed Self-Transmissible Plasmids Widely Distributed among Different Environments in Japan. Appl. Environ. Microbiol., e0111422.

36. Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2011). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat. Methods *9*, 173–175.

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

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

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

40. Giraldo, R. (2003). Common domains in the initiators of DNA replication in*Bacteria, Archaea*and*Eukarya*: combined structural, functional and phylogenetic perspectives. FEMS Microbiol. Rev. *26*, 533–554.

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

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

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

44. Haines, A.S., Akhtar, P., Stephens, E.R., Jones, K., Thomas, C.M., Perkins, C.D., Williams, J.R., Day, M.J., and Fry, J.C. (2006). Plasmids from freshwater environments capable of IncQ retrotransfer are diverse and include pQKH54, a new IncP-1 subgroup archetype. Microbiology *152*, 2689–2701.

45. Lam, M.M.C., and Hamidian, M. (2024). Examining the role of Acinetobacter

baumannii plasmid types in disseminating antimicrobial resistance. NPJ Antimicrob Resist *2*, 1.

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

47. del Solar, G., Giraldo, R., Ruiz-Echevarría, M.J., Espinosa, M., and Díaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. Microbiol. Mol. Biol. Rev. *62*, 434–464.

48. Ilyina, T.V., and Koonin, E.V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Res. *20*, 3279–3285.

49. Koonin, E.V., Kuhn, J.H., Dolja, V.V., and Krupovic, M. (2024). Megataxonomy and global ecology of the virosphere. ISME J. *18*. https://doi.org/10.1093/ismejo/wrad042.

50. Neri, U., Wolf, Y.I., Roux, S., Camargo, A.P., Lee, B., Kazlauskas, D., Chen, I.M., Ivanova, N., Zeigler Allen, L., Paez-Espino, D., et al. (2022). Expansion of the global RNA virome reveals diverse clades of bacteriophages. Cell *185*, 4023–4037.e18.

51. Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. *36*, 996–1004.

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

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

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

55. plasmidSPAdes – Center for Algorithmic Biotechnology https://cab.spbu.ru/software/plasmid-spades/.

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

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

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

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

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

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

62. Lobry, J.R. (1996). Origin of Replication of *Mycoplasma genitalium*. Preprint, https://doi.org/10.1126/science.272.5262.745 https://doi.org/10.1126/science.272.5262.745.

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

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

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

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

67. Galata, V., Fehlmann, T., Backes, C., and Keller, A. (2018). PLSDB: a resource of complete bacterial plasmids. Nucleic Acids Res. https://doi.org/10.1093/nar/gky1050.

68. Schmartz, G.P., Hartung, A., Hirsch, P., Kern, F., Fehlmann, T., Müller, R., and Keller, A. (2022). PLSDB: advancing a comprehensive database of bacterial plasmids.

Nucleic Acids Res. *50*, D273–D278.

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

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

71. Bayliss, S.C., Thorpe, H.A., Coyle, N.M., Sheppard, S.K., and Feil, E.J. (2019). PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. Gigascience *8*. https://doi.org/10.1093/gigascience/giz119.

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

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

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

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

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

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

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

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

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

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

82. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

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

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

85. Camargo, A.P., Call, L., Roux, S., Nayfach, S., Huntemann, M., Palaniappan, K., Ratner, A., Chu, K., Mukherjeep, S., Reddy, T.B.K., et al. (2024). IMG/PR: a database of plasmids from genomes and metagenomes with rich annotations and metadata. Nucleic Acids Res. *52*, D164–D173.

86. Pansegrau, W., Lanka, E., Barth, P.T., Figurski, D.H., Guiney, D.G., Haas, D., Helinski, D.R., Schwab, H., Stanisich, V.A., and Thomas, C.M. (1994). Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. J. Mol. Biol. *239*, 623–663.

87. Shintani, M., Suzuki, H., Nojiri, H., and Suzuki, M. (2022). Precise classification of antimicrobial resistance-associated IncP-2 megaplasmids for molecular epidemiological studies on Pseudomonas species. J. Antimicrob. Chemother. *77*, 1203–1205.

88. Llanes, C., Gabant, P., Couturier, M., Bayer, L., and Plesiat, P. (1996). Molecular analysis of the replication elements of the broad-host-range RepA/C replicon. Plasmid *36*, 26–35.

89. Fricke, W.F., Welch, T.J., McDermott, P.F., Mammel, M.K., LeClerc, J.E., White, D.G., Cebula, T.A., and Ravel, J. (2009). Comparative genomics of the IncA/C multidrug resistance plasmid family. J. Bacteriol. *191*, 4750–4757.

90. Scholz, P., Haring, V., Wittmann-Liebold, B., Ashman, K., Bagdasarian, M., and Scherzinger, E. (1989). Complete nucleotide sequence and gene organization of the broad-host-range plasmid RSF1010. Gene *75*, 271–288.

91. Maeda, K., Nojiri, H., Shintani, M., Yoshida, T., Habe, H., and Omori, T. (2003). Complete nucleotide sequence of carbazole/dioxin-degrading plasmid pCAR1 in Pseudomonas resinovorans strain CA10 indicates its mosaicity and the presence of large catabolic transposon Tn4676. J. Mol. Biol. *326*, 21–33.

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

93. Takahashi, Y., Shintani, M., Yamane, H., and Nojiri, H. (2009). The complete nucleotide sequence of pCAR2: pCAR2 and pCAR1 were structurally identical IncP-7 carbazole degradative plasmids. Biosci. Biotechnol. Biochem. *73*, 744–746.

94. Greated, A., Lambertsen, L., Williams, P.A., and Thomas, C.M. (2002). Complete sequence of the IncP-9 TOL plasmid pWW0 from Pseudomonas putida. Preprint, https://doi.org/10.1046/j.1462-2920.2002.00305.x https://doi.org/10.1046/j.1462-2920.2002.00305.x.

95. Sota, M., Yano, H., Ono, A., Miyazaki, R., Ishii, H., Genka, H., Top, E.M., and Tsuda, M. (2006). Genomic and functional analysis of the IncP-9 naphthalene-catabolic plasmid NAH7 and its transposon Tn4655 suggests catabolic gene spread by a tyrosine recombinase. J. Bacteriol. *188*, 4057–4067.

Table 1. Lists of representative plasmids in the PIncs.

PInc(s) (classical Inc name)	Representative plasmid [accession no.]	MOB MPF[a]	No. of plasmids in PLSDB	Numbers of WH	Clade	References
PInc-1 **(IncP-1, IncP)**	RK2/RP4 [BN000925]	MOB_P MPF_T	321	2	F	[86]
PInc-2 **(IncP-2)**	**Rms139** **[LC653116]**	MOB_P MPF_I	87	2	E	[13,87]
PInc-3 **(IncP-3, IncA/C)**	pRA1 (IncA) [FJ705807] pR55 (IncC) [JQ010984]	MOB_H MPF_F	921	2	E	[88,89]
PInc-4 **(IncP-4, IncQ)**	RSF1010 [M28829]	MOB_Q -	695	2	F	[90]

PInc-5 **(IncP-5)**	**Rms163** **[LC685027]**	NA NA	2	1	B	[14]
PInc-6 **(IncP-6, IncG)**	Rms149 [AJ877225]	MOB_P -	119	NA	-	[43]
PInc-7 **(IncP-7)**	pCAR1 [AB088420]	MOB_H MPF_F	23	2	H	[91–93]
PInc-9 **(IncP-9)**	pWW0 [AJ344068] NAH7 [AB237655] **Rsu2** [**LC685593]**	MOB_F MPF_T	34	2	A	[17,94,95]
PInc-10 **(IncP-10)**	R91-5 [X54695[b]] pPAB546 [MN433456]	NA NA	49	1	C	[22,23]

PInc-11 **(IncP-11)**	**RP1-1** **[LC700336]**	MOB_c NA	23	1	C	[18]
PInc-12 **(IncP-12)**	**R716** [**LC685026**]	NA NA	9	2	F	[19]
- **(IncP-13)**	**pMG26**[c] **[LC685025]**	MOB_F MPF_G	NA	NA	-	[20,21]
PInc-15 **(PromA)**	pSN1104-11 [AP018707]	MOB_P MPF_T	42	2	G	[35]
PInc-16	pSN1216-29 [AP018710]	MOB_P MPF_T	24	2	E	[30]
PInc-17	pQBR103 [AM235768]	NA NA MOB_F MPF_F	11	2	E	[31]

PInc-18 **(Inc_{pSTY})**	pSTY [NC_022739] pPT23-C1_1 [deposited]	MOB_F MPF_F	15	2	H	[32]

[a]Classification of MOB classes and MPF types was identified by the MOB-suite [87] with criteria of >=90% identity and >=60% coverage. "-" indicates that the genes involved in conjugation have not been detected, whereas "NA" indicates that the classes and types of MOB or MPF were not identified although the plasmid was self-transmissible. [b]Partial DNA sequences are available. [c]pMG26 was shown to be not a plasmid but an integrative and conjugative element in this study.

Figure 1. Predicted three-dimensional structures of RIPs of plasmids in PIncs.
(A) Predicted structures of RIPs of PInc-1 (RepP-1A (=TrfA33) of RK2), PInc-2 (RepP-2A of Rms139), PInc-3 (RepP-3A of pRA1), PInc-4 (RepP-4A (=RepC) of pRSF1010), PInc-7 (RepP-7A of pCAR1), PInc-12 (RepP-12A of R716), PInc-15 (RepP-15A of pSN1104-11), PInc-16 (RepP-16A of pSB1216-29), PInc-17 (RepP-17A of pQBR103), and PInc-18 (RepP-18A of pSTY). The crystal structure (PDB: 2NRA) of Pi protein of IncX plasmid, R6K, interacted with iteron-containing dsDNA is shown as reference, of which structure is elucidated by X-ray crystallography. The WH domains of Pi are shown in pink and blue (WH1: 29-139 aa; WH2: 164-263 aa). (B) Predicted structures of RIPs of PInc-5 (RepP-5A of Rms163), PInc-9 (RepP-9A of pWW0), PInc-10 (RepP-10A of pPAB546), and PInc-11 (RepP-11A of RP1-1). The crystal structure (subunit B in PDB: 4PT7) of Rep19A proteins of *Staphylococcus* plasmid, pCA-347, is shown as reference, of which structure is elucidated by X-ray crystallography. The WH domain of Rep19A is shown in blue (WH: 29-111 aa).

Figure 2. Features of plasmids of PIncs in the PLSDB.

Distribution of (A) plasmid sizes (kb) and (B) G+C contents (%), and (C) ratio of core and accessory genes of each PIncs plasmids in the PLSDB. (D) Relationship of each PIncs plasmids to other plasmid replicons (left) and ARGs (right) in the PLSDB. “rep_cluster” contains different types of replicons identified by MOB-typer. ’N.D.' indicates not detected. A sankey diagram was generated using the SankeyWidget library in Python through in-house programming.

Figure 3. A large-scale phylogeny of WH RIPs.

(A) A maximum-likelihood tree of WH RIPs (4,769 leaves; nonredundant at ≥50% protein identity) based on the conserved region (150 positions of the protein sequence alignment). These leaves are associated with 34,724 PLSDB plasmids, 95,057 UniParc sequences, and 135,545 IMG/PR plasmids according to sequence similarity of RIPs (basically ≥50% identity). The defined 8 clades and 42 subclades are shown. The tree was rooted by the split of sWH RIPs (clade A-D) and dWH RIPs (clade E-H). Rings represent information of associated sequences for each leaf as follows: (from the outermost to innermost) groups assigned by iterative collection process of RIPs; Pfam, taxonomy, and environmental assignment proportion; the number of associated IMG/PR, UniParc, and PLSDB sequences; amino acid length, N-ARSC (nitrogen atoms per residue side chain), C-ARSC (carbon atoms per residue side chain), and %GC of associated RIPs from UniParc and PLSDB; nucleotide length and %GC of associated PLSDB plasmids; %MOB and %MPF types of PLSDB plasmids; the number of

AMR, defence, and anti-defence systems for PLSDB plasmids; the number of multi-replicons, which means the number of multiple WH RIPs in a plasmid, calculated for each RIP of PLSDB plasmids; (B) The clade and subclade names, support values, and number of leaves, and the associated replicon types identified with PlasmidFinder, MOB-suite, and AcinetobacterPlasmidTyping are shown. In addition, crystal structure resolved (*) and predicted RIPs (**) are shown.

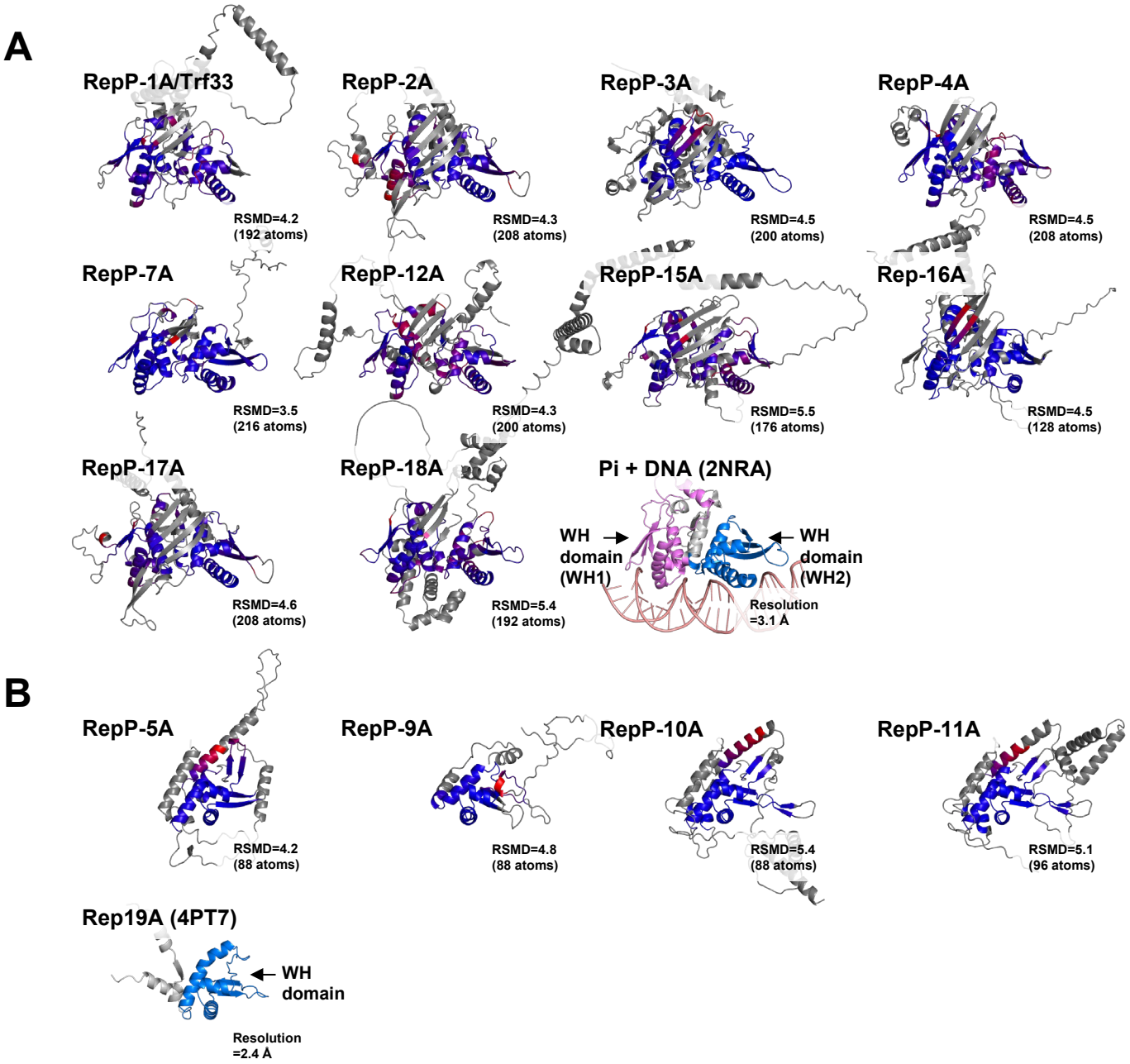
A
RepP-1A/Trf33
RSMD=4.2
(192 atoms)
RepP-2A
RSMD=4.3
(208 atoms)
RepP-3A
RSMD=4.5
(200 atoms)
RepP-4A
RSMD=4.5
(208 atoms)
RepP-7A
RSMD=3.5
(216 atoms)
RepP-12A
RSMD=4.3
(200 atoms)
RepP-15A
RSMD=5.5
(176 atoms)
Rep-16A
RSMD=4.5
(128 atoms)
RepP-17A
RSMD=4.6
(208 atoms)
RepP-18A
RSMD=5.4
(192 atoms)
Pi + DNA (2NRA)
WH domain (WH1)
WH domain (WH2)
Resolution =3.1 Å
B
RepP-5A
RSMD=4.2
(88 atoms)
RepP-9A
RSMD=4.8
(88 atoms)
RepP-10A
RSMD=5.4
(88 atoms)
RepP-11A
RSMD=5.1
(96 atoms)
Rep19A (4PT7)
WH domain
Resolution =2.4 Å

Fig. 1

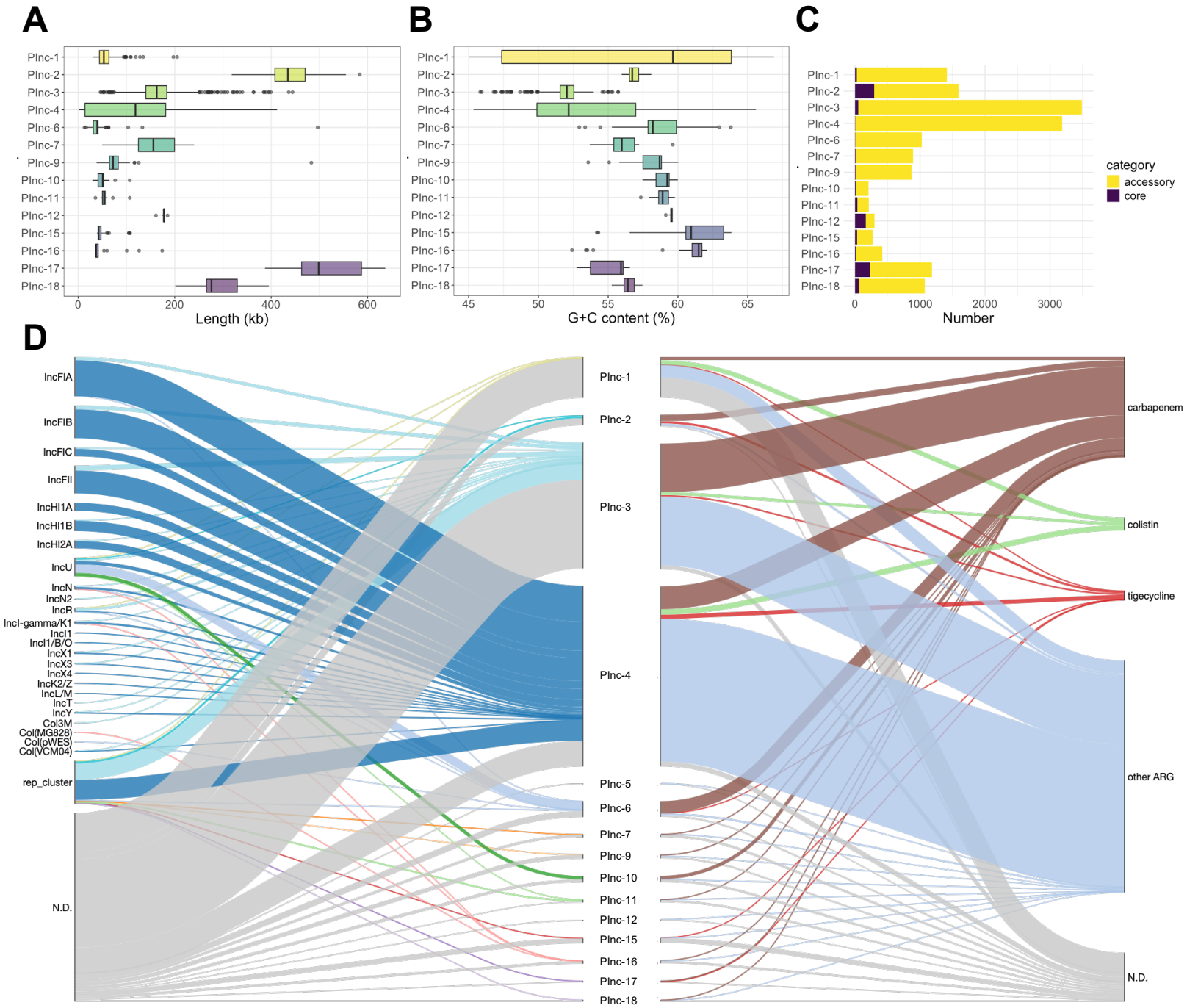
A
PInc-1
PInc-2
PInc-3
PInc-4
PInc-6
PInc-7
PInc-9
PInc-10
PInc-11
PInc-12
PInc-15
PInc-16
PInc-17
PInc-18
0
200
400
600
Length (kb)
B
45
50
55
60
65
G+C content (%)
C
0
1000
2000
3000
Number
category
accessory
core
D
IncFIA
IncFIB
IncFIC
IncFII
IncHI1A
IncHI1B
IncHI2A
IncU
IncN
IncN2
IncR
IncI-gamma/K1
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. 2

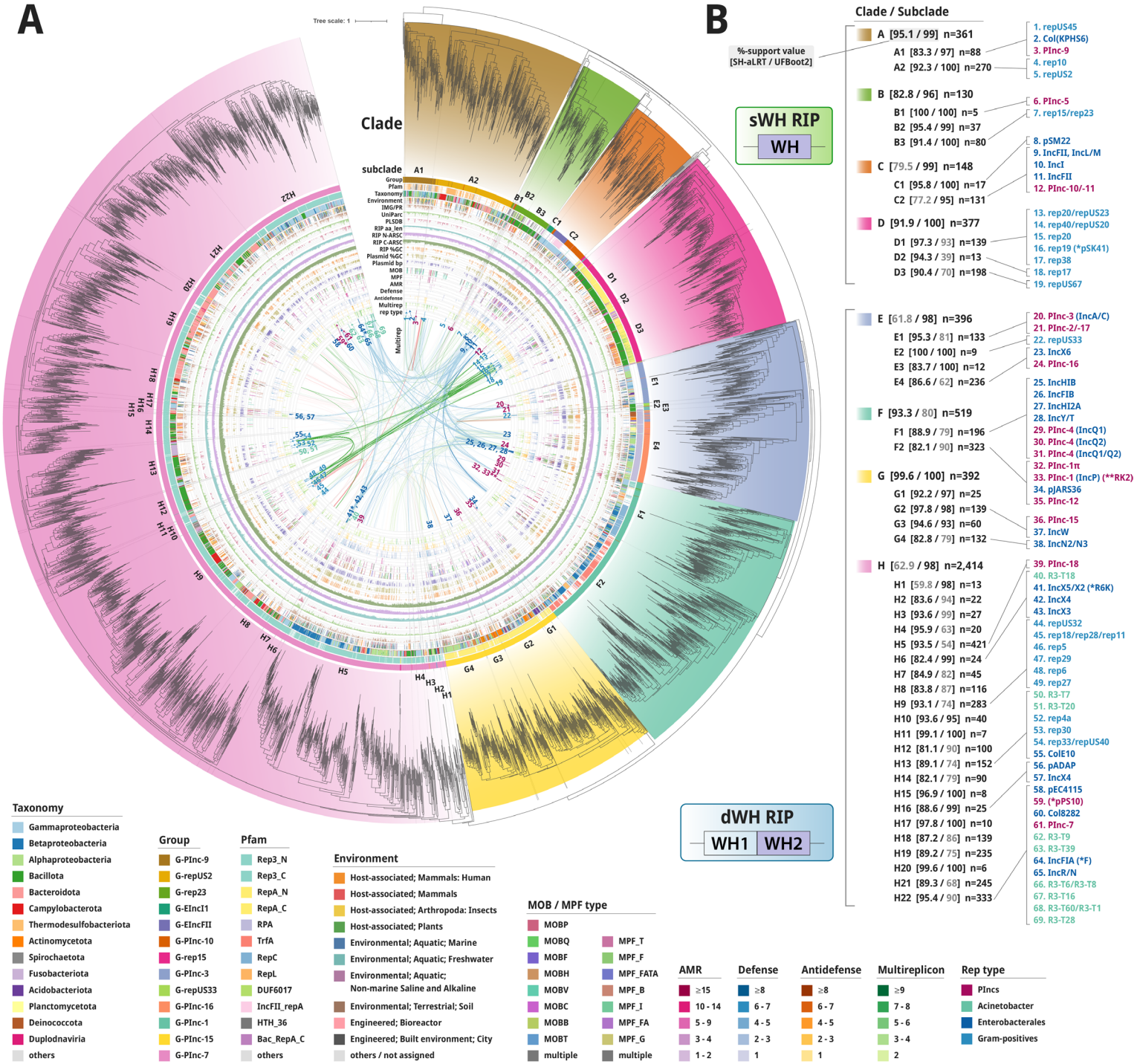
A
B
Tree scale: 1
Clade
subclade
Group
Pfam
Taxonomy
Environment
IMG/PR
UniParc
PLSDB
RIP aa len
RIP N-ARSC
RIP C-ARSC
RIP %GC
Plasmid %GC
Plasmid bp
MOB
MPF
AMR
Defense
Antidefense
Multirep
rep type
Multirep
%-support value
[SH-aLRT / UFBoot2]
sWH RIP
WH
dWH RIP
WH1
WH2
Clade / Subclade
A [95.1 / 99] n=361
A1 [83.3 / 97] n=88
A2 [92.3 / 100] n=270
B [82.8 / 96] n=130
B1 [100 / 100] n=5
B2 [95.4 / 99] n=37
B3 [91.4 / 100] n=80
C [79.5 / 99] n=148
C1 [95.8 / 100] n=17
C2 [77.2 / 95] n=131
D [91.9 / 100] n=377
D1 [97.3 / 93] n=139
D2 [94.3 / 39] n=13
D3 [90.4 / 70] n=198
E [61.8 / 98] n=396
E1 [95.3 / 81] n=133
E2 [100 / 100] n=9
E3 [83.7 / 100] n=12
E4 [86.6 / 62] n=236
F [93.3 / 80] n=519
F1 [88.9 / 79] n=196
F2 [82.1 / 90] n=323
G [99.6 / 100] n=392
G1 [92.2 / 97] n=25
G2 [97.8 / 98] n=139
G3 [94.6 / 93] n=60
G4 [82.8 / 79] n=132
H [62.9 / 98] n=2,414
H1 [59.8 / 98] n=13
H2 [83.6 / 94] n=22
H3 [93.6 / 99] n=27
H4 [95.9 / 63] n=20
H5 [93.5 / 54] n=421
H6 [82.4 / 99] n=24
H7 [84.9 / 82] n=45
H8 [83.8 / 87] n=116
H9 [93.1 / 74] n=283
H10 [93.6 / 95] n=40
H11 [99.1 / 100] n=7
H12 [81.1 / 90] n=100
H13 [89.1 / 74] n=152
H14 [82.1 / 79] n=90
H15 [96.9 / 100] n=8
H16 [88.6 / 99] n=25
H17 [97.8 / 100] n=10
H18 [87.2 / 86] n=139
H19 [89.2 / 75] n=235
H20 [99.6 / 100] n=6
H21 [89.3 / 68] n=245
H22 [95.4 / 90] n=333
1. repUS45
2. Col(KPHS6)
3. PInc-9
4. rep10
5. repUS2
6. PInc-5
7. rep15/rep23
8. pSM22
9. IncFII, IncL/M
10. IncI
11. IncFII
12. PInc-10/-11
13. rep20/repUS23
14. rep40/repUS20
15. rep20
16. rep19 (*pSK41)
17. rep38
18. rep17
19. repUS67
20. PInc-3 (IncA/C)
21. PInc-2/-17
22. repUS33
23. IncX6
24. PInc-16
25. IncHIB
26. IncFIB
27. IncHI2A
28. IncY/T
29. PInc-4 (IncQ1)
30. PInc-4 (IncQ2)
31. PInc-4 (IncQ1/Q2)
32. PInc-1π
33. PInc-1 (IncP) (**RK2)
34. pJARS36
35. PInc-12
36. PInc-15
37. IncW
38. IncN2/N3
39. PInc-18
40. R3-T18
41. IncX5/X2 (*R6K)
42. IncX4
43. IncX3
44. repUS32
45. rep18/rep28/rep11
46. rep5
47. rep29
48. rep6
49. rep27
50. R3-T7
51. R3-T20
52. rep4a
53. rep30
54. rep33/repUS40
55. ColE10
56. pADAP
57. IncX4
58. pEC4115
59. (*pPS10)
60. Col8282
61. PInc-7
62. R3-T9
63. R3-T39
64. IncFIA (*F)
65. IncR/N
66. R3-T6/R3-T8
67. R3-T16
68. R3-T60/R3-T1
69. R3-T28
Taxonomy
Gammaproteobacteria
Betaproteobacteria
Alphaproteobacteria
Bacillota
Bacteroidota
Campylobacterota
Thermodesulfobacteriota
Actinomycetota
Spirochaetota
Fusobacteriota
Acidobacteriota
Planctomycetota
Deinococcota
Duplodnaviria
others
Group
G-PInc-9
G-repUS2
G-rep23
G-EIncI1
G-EIncFII
G-PInc-10
G-rep15
G-PInc-3
G-repUS33
G-PInc-16
G-PInc-1
G-PInc-15
G-PInc-7
Pfam
Rep3_N
Rep3_C
RepA_N
RepA_C
RPA
TrfA
RepC
RepL
DUF6017
IncFII_repA
HTH_36
Bac_RepA_C
others
Environment
Host-associated; Mammals: Human
Host-associated; Mammals
Host-associated; Arthropoda: Insects
Host-associated; Plants
Environmental; Aquatic; Marine
Environmental; Aquatic; Freshwater
Environmental; Aquatic; Non-marine Saline and Alkaline
Environmental; Terrestrial; Soil
Engineered; Bioreactor
Engineered; Built environment; City
others / not assigned
MOB / MPF type
MOBP
MOBQ
MOBF
MOBH
MOBV
MOBC
MOBB
MOBT
multiple
MPF_T
MPF_F
MPF_FATA
MPF_B
MPF_I
MPF_FA
MPF_G
multiple
AMR
≥15
10 - 14
5 - 9
3 - 4
1 - 2
Defense
≥8
6 - 7
4 - 5
2 - 3
1
Antidefense
≥8
6 - 7
4 - 5
2 - 3
1
Multireplicon
≥9
7 - 8
5 - 6
3 - 4
2
Rep type
PIncs
Acinetobacter
Enterobacterales
Gram-positives

Fig. 3