Mini Review

Reconsidering plasmid maintenance factors for computational plasmid design

Hirokazu Yano⁎, Masaki Shintani b,c, Masaru Tomita d,e, Haruo Suzuki d,e,⁎⁎, Taku Oshima f

a Graduate School of Life Sciences, Tohoku University, 2-1-1, Katahira, Aoba-ku, Sendai 980-8577, Japan
b Department of Engineering, Graduate School of Integrated Science and Technology, Shizuoka University, 3-5-1, Hamamatsu 432-8561, Japan
c Department of Bioscience, Graduate School of Science and Technology, Shizuoka University, 3-5-1, Hamamatsu 432-8561, Japan
d Institute for Advanced Biosciences, Keio University, 14-1, Baba-cho, Tsuruoka, Yamagata 997-0035, Japan
e Faculty of Environment and Information Studies, Keio University, 5322, Endo, Fujisawa, Kanagawa 252-0882, Japan
f Department of Biotechnology, Toyama Prefectural University, 5180, Kurokawa, Itami, Toyama 939-0398, Japan

A R T I C L E  I N F O

Article history:
Received 23 September 2018
Received in revised form 8 December 2018
Accepted 9 December 2018
Available online 15 December 2018

Keywords:
Bioinformatics
Genome design
Vector design
Synthetic biology
Plasmid persistence
Host-parasite coevolution

A B S T R A C T

Plasmids are genetic parasites of microorganisms. The genomes of naturally occurring plasmids are expected to be polished via natural selection to achieve long-term persistence in the microbial cell population. However, plasmid genomes are extremely diverse, and the rules governing plasmid genomes are not fully understood. Therefore, computationally designing plasmid genomes optimized for model and nonmodel organisms remains challenging. Here, we summarize current knowledge of the plasmid genome organization and the factors that can affect plasmid persistence, with the aim of constructing synthetic plasmids for use in gram-negative bacteria. Then, we introduce publicly available resources, plasmid data, and bioinformatics tools that are useful for computational plasmid design.

© 2018 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Contents

1. Introduction ...............................................................71
2. Key Factors in Plasmid Design .......................................................71
   2.1. Plasmid Gene Content ........................................................71
   2.1.1. Defining the Plasmid Core ..................................................71
   2.1.2. Functional Modules Comprising a Plasmid ........................................71
   2.1.3. Testing the Functionality of Functional Modules ........................................75
   2.1.4. Selection Markers .....................................................75
   2.2. Interactions with the Host ......................................................75
   2.2.1. Host Factors ........................................................75
   2.2.2. Fitness Cost Imposed by Plasmids .......................................................75
   2.3. Constraints in the Genome .....................................................75
   2.3.1. Size Constraint ........................................................75
   2.3.2. G + C Content ........................................................76
   2.3.3. Oligonucleotide Composition .......................................................76
   2.3.4. Codon Usage ........................................................76
   2.3.5. Gene Direction .......................................................77
3. Publicly Available Resources ........................................................77
   3.1. Plasmid Sequence Data .......................................................77
   3.2. Bioinformatics Tools .......................................................77
4. Concluding Remarks ............................................................77

⁎ Corresponding author.
⁎⁎ Corresponding author at: Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan.
E-mail addresses: yano.hirokazu@ige.tohoku.ac.jp (H. Yano), haruo@sfc.keio.ac.jp (H. Suzuki).

https://doi.org/10.1016/j.csbj.2018.12.001
2001-0370/© 2018 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
1. Introduction

Plasmids are autonomously replicating DNA molecules present in microorganisms. Plasmids are also known to be mobile genetic elements that can be horizontally transferred among different organisms [1,2]. Plasmids can be considered as genetic parasites in the sense that their reproduction depends to some extent on their host and that they do not necessarily share the fate of a specific cell lineage, as they are horizontally transmissible. Plasmids have been used as primal genetic tools for exogenous DNA expression and microbial metabolic engineering. The importance of plasmid vectors has increased in recent years [3]. Currently, the plasmid genome is difficult to design computationally because the elements contained within plasmids are not conserved across plasmid groups. Additionally, a number of factors affect plasmid persistence. Understanding the key factors affecting replication and stable maintenance of plasmids in a host cell population is essential to control plasmids as synthetic vectors. Conversely, construction of synthetic vectors based on our knowledge and testing its persistence in a model host could indicate how far we have to go to understand undiscovered plasmid maintenance factors. If designed plasmids are stably maintained, it follows that the selected elements (genes, intergenic regions, etc.) play a positive role in plasmid persistence.

In this review, we summarize current knowledge of the key factors that affect plasmid persistence and then introduce publicly available resources (plasmid data and bioinformatics tools) potentially useful for designing synthetic plasmids, aiming at their use in exogenous DNA expression and microbial metabolic engineering.

2. Key Factors in Plasmid Design

2.1. Functional Modules Comprising a Plasmid

Based on recent progress in plasmid biology and bioinformatics, we consider three factors that should be taken into account to design plasmids as synthetic vectors. These factors are described in detail in the subsequent sections.

2.1.1. Defining the Plasmid Core

Plasmids show gene content variations, even within the same Inc. group [17]. Thus, plasmids are likely to experience gene gain and loss over evolutionary time [18,19]. A comparative analysis of closely related taxa can categorize a genome into two parts: (i) “core” genes conserved in all members within a defined group (e.g., bacterial species, Inc. group, etc.), and (ii) “noncore” genes absent in some members within the group. Being a core gene does not necessarily mean that the gene positively contributes to plasmid maintenance in particular hosts, but suggests that the gene sets have co-evolved together since the divergence from the most recent common ancestor. The long-term co-evolution of core genes can result in the formation of an operon with a coordinated regulatory system that balances the efficiency of horizontal and vertical transmissions [20–22]. These core genes may be linked together upon construction of a vector. A recent analysis of recombination tracts in the plasmid core genome highlighted a block of evolutionarily linked genes [23]. These findings also suggest that the plasmid core undergoes recombinational allelic exchange within the group at an evolutionary time scale.

Core and noncore genes can be identified by homologous gene clustering for a defined plasmid group, e.g., using all-against-all protein sequence comparisons with BLASTP [24]. We previously found that homolog clusters specific to each of the six Inc. groups (F, H, I, N, P-1, and W) (Table 1) were involved in plasmid replication, partition, and transfer [17]. Based on the BLASTP (E-value <e−5) comparison, replication initiation (Rep) proteins for the six Inc. groups (RepB and RepE for IncF, RepA for IncFI, RepHIA for IncH, RepZ for IncI, RepA for IncN, TrfA for IncP-1, and RepA for IncW) formed distinct homolog clusters (exceptions were RepB and RepHIA, which formed a single homolog cluster) and were conserved in all members within each of the Inc. groups.

In the IncP-1 group, TrfA (replication initiation protein), KlcA (antirestriction protein), KlcC (KorC transcriptional repressor), KleE (stable inheritance protein), KfrA, IncC (ParA homolog), and KorB (ParB homolog) were conserved in all 22 plasmids analyzed [17] and thus deemed as the plasmid core. Surprisingly, a homolog cluster for DNA transfer (tra, trb), postsegregational cell killing (parDE, relBE), multimer resolution (parA), and regulatory protein (korA), which was shown to support plasmid maintenance ([25]), was not identified as the plasmid core. The absence of genes suggests lineage-specific gene loss or nonorthologous gene displacement during plasmid evolution in nature [18,19].

2.1.2. Functional Modules Comprising a Plasmid

Gene products, which contribute to plasmid maintenance in bacterial hosts, require cis-acting sites to elicit their functions. In this review, a functional module is defined as a pair of gene products and its acting site on a plasmid. Each functional module often contains its own regulatory function. In such cases, the elements of each functional module should not be separated upon construction of a synthetic plasmid. Below, we briefly describe the features of representative functional modules comprising a plasmid, i.e., replication module, partition module, toxin-antitoxin module, multimer resolution module, DNA transfer module, and antirestriction module. Plasmid genomes are often considered an assembly of these functional modules (Fig. 1). Plasmid functional modules are potential sources for biological parts for synthetic biology projects, such as BioBrick [26] and SEVA [27].
which are produced from near the oriV [28]. RNA II is converted to primer RNA (thus acts as an initiator of replication), whereas RNA I and Rop protein cooperatively inhibit RNA II maturation (thus act as inhibitors of replication). The copy number of ColE1-type plasmids is maintained at around 10–15 copies/cell [9]. This type of replicon has been used as a cloning vector, including pUC and pET vectors (Table 1). For pUC vectors, deletion of the Rop protein gene and a putative mating assays.

**Table 1**

| Incompatibilitya | Representative plasmid (original host)b | Accession numberc | RIPd | MOBd | MPPd | Host rangef | Reference |
|------------------|-----------------------------------------|-------------------|------|------|------|-------------|-----------|
| Inc groups       |                                         |                   |      |      |      |             |           |
| A/C1 (= IncA)    | R1 (Aeromonas hydrophila)               | NC_012885         | RepA | MOB4 | MPP2 | Gammaproteobacteria | [151]     |
| A/C2 (= IncC)    | pMH1760 (Klebsiella pneumoniae)         | KF967462          | RepA | MOB4 | MPP2 | Gammaproteobacteria | [152]     |
| IncB/D           | R3521 (Escherichia coli)               | GU256641          | RepA | NA   | NA   | NA          | [153]     |
| IncD             | R711b (Providencia)                   | NA                | NA   | NA   | NA   | NA          |           |
| IncF             | F (Escherichia coli)                  | AP001918          | RepE (for RepFIA replicon) | MOB4 | MPP2 | Entero bacteriaceae, Yersiniaceae, Escherichia, Salmonella, Proteus | [155]     |
| IncH             | R1 (Salmonella enterica)              | KY749237          | RepA | MOB4 | MPP2 | Escherichia, Salmonella, Proteobacteria | [156]     |
| IncG/U (= IncP-6) | RA3 (Pseudomonas aeruginosa)          | DQ401103          | RepA | MOB4 | MPP2 | Enterobacteriaceae, Yersiniaceae, Erwinia, Escherichia, Salmonella, Shigella | [159]     |
| IncH             | R27 (Salmonella typhi)                | AF250878          | RepH1A, RepH1B | MOB4 | MPP2 | Enterobacteriaceae, Yersiniaceae, Erwinia, Escherichia, Salmonella, Shigella | [159]     |
| IncI             | R64 (Salmonella enterica)             | AP001547          | RepZ | MOB4 | MPP2 | Proteobacteria | [160]     |
| IncJ             | R391 (Providencia rettgeri)           | AY090559          | –    | MOB4 | MPP2 | NA          | [161]     |
| IncK             | R387 (Shigella flexneri)              | NCTC50022         | RepA | MOB4 | MPP2 | NA          | [162]     |
| IncL             | R471 (Serratia marcescens)            | KM406489          | RepA | MOB4 | MPP2 | Protein | [163,164] |
| IncM             | R69 (Salmonella enterica)             | KM406488          | RepA | MOB4 | MPP2 | Proteobacteria | [165]     |
| IncN             | N3 (Escherichia coli)                 | NC_015599         | RepA | MOB4 | MPP2 | Proteobacteria | [166]     |
| IncP (= IncP-1)  | RK2 (Pseudomonas aeruginosa)          | BN000925          | TriA | MOB4 | MPP2 | Proteobacteria | [167]     |
| IncQ             | R5F1010 (Escherichia coli)            | M28829            | RepABC | MOB4 | MPP2 | Proteobacteria | [168]     |
| IncR             | pP1780 (Klebsiella pneumoniae)        | JX424614          | RepB | –    | –    | Klebsiella | [169]     |
| IncS (= IncH2)   | R478 (Serratia marcescens)            | BX664015          | RepHA | MOB4 | MPP2 | Serratia | [170]     |
| IncT             | R753 (Escherichia coli)               | NCTC50521 (planned) | NA | NA | NA | NA | [172]     |
| IncW             | R388 (Escherichia coli)               | NC_028464         | RepA | MOB4 | MPP2 | Salmonella, Erwinia, Escherichia, Providencia | [173]     |
| IncX             | R6K (Escherichia coli)                | NCTC50005         | π    | MOB4 | MPP2 | Enterobacteriaceae | [174]     |
| IncY             | P1 (Escherichia virus)                | AF234172 (phage P1 mod749::IS5 c1–100) | RepA | – | – | Enterobacteriaceae | [175,176] |
| IncZ             | pMCR-1-P3 (Escherichia coli)          | KX850944          | RepA | NA | NA | Klebsiella | [162]     |
| PromA            | pMRA020 (Methylbacterium radiotolerans) | NC_010509       | RepA | MOB4 | MPP2 | Proteobacteria | [177–179] |
| IncP-9           | NAH7 (Pseudomonas putida)             | AF078924 (partial) | Rep | MOB4 | MPP2 | Pseudomonas, Escherichia | [180,181] |

**Representative cloning vectors**

- Not assigned pUC18/19, pET, pBluescript L09136 pMB1 copy-up type, pMB1 type, pMB1 copy-up type – – Escherichia [175] Merck Mustard [182]
- Not assigned pACYC184 XO6403 p15A type pMR1 Rep – – Escherichia [183] – – – [184]
- Not assigned pBR1MCS (Bordetella bronchiseptica) NC_025015 pBR1 Rep MOB4 – – – [184]
- Not assigned pSC101 (Salmonella enterica) NC_002056 RepA MOB4 – – Salmonella, Escherichia [185]
- Not assigned pB1C1 (Rhizobium etli) KYO31728 p42b RepC – – Rhizobiales [44]
- Not assigned pME8041 (Pseudomonas sp.) AF118812 pVS1 RepA – – Proteobacteria [65]

---

a Several Inc groups are identical; e.g. IncG = IncU.
b Representative plasmids are listed based on Lawley et al. [186].
c Accession numbers in National Center for Biotechnology Information (NCBI) GeneBank and RefSeq (https://www.ncbi.nlm.nih.gov/Sequin/acc.html) and Wellcome Sanger Institute (prefix “NCBI” https://www.sanger.ac.uk/resources/downloads/plasmids/).
d Names of replication initiation protein (RIP). “NA” indicates that the nucleotide sequences of the plasmid are not available.
e Classification of MOB classes and MPP types is based on Smillie et al. [132] and Guglielmini et al. [187]. “-” indicates that the genes involved in conjugation have not been detected, whereas “NA” indicates that the nucleotide sequences of the plasmid are not available.
f Plasmid host range determined based on genome sequencing projects (hosts in which a plasmid has been found) and/or filter mating assays.
g Original hosts are unknown because exogenous plasmid capturing was used.

† Plasmid host range determined based on genome sequencing projects (hosts in which a plasmid has been found) and/or filter mating assays.

H. Yano et al. / Computational and Structural Biotechnology Journal 17 (2019) 70–81
Streptomyces theta-type replication is unidirectional (exceptions include the linear other modules assist in plasmid maintenance. IncP-1 Fig. 1. division. These mechanisms are equivalent to the function of the spindle oriV + T-rich region within a replicon that has evolved in species closely related to a model host. Transcriptional regulator or NAPs (H-NS homologs) for plasmid genes could reduce the fitness cost imposed by the plasmid.

The G + C content of the plasmid should match that of the host. Highly expressed essential genes should be on leading strands.

Table 2
| Factors | Notes (what should be considered) |
|---------|----------------------------------|
| 1. Plasmid gene content | Include a set of plasmid core genes. Include selection marker or a toxin-antitoxin system to prevent generation of plasmid-free cells. Include cis-elements, such as centromere-like site and resolution site. |
| 2. Interaction with host | Select a basic replicon that has evolved in species closely related to a model host. Transcriptional regulator or NAPs (H-NS homologs) for plasmid genes could reduce the fitness cost imposed by the plasmid. |
| 3. Constraints in genome | The G + C content of the plasmid should match that of the host. Highly expressed essential genes should be on leading strands. |

(500–700 copies/cell) [29]. The replication modules of so-called iteron-containing plasmids, e.g., RK2 and R6K (Table 1), consist of a replication initiation protein (Rep protein) gene and oriV, which are in general located next to each other on the plasmid (Fig. 2A). oriV contains a Rep protein-binding region (iterons), host DnaA-binding region (DnaA-boxes), and DNA unwinding elements (DUE), which are motifs in an A + T-rich region within oriV [8]. Rep proteins act as both initiators and inhibitors of replication [8,28,30]. Purified Rep proteins are mostly dimeric, whereas only monomeric Rep protein is active in unwinding DUE (Fig. 2B). DnaB helicase is loaded onto unwind DUE via either a host DnaA-dependent or -independent manner. Rep protein can also bind a specific strand of unwound DUE and assists replisome assembly on one strand via direct interaction with 3'-clamp, leading to unidirectional replication [31]. Dimeric Rep proteins prevent oriV melting by pairing iterons in a phenomenon called handcrafting (Fig. 2C) [32–34]. An increased monomer to dimer ratio dissociates the paired iterons [32,33].

A Rep protein mutant of R6K n protein (pir-116 allele) [35] lacks replication inhibition activity (unable to form dimer) and has been used to increase the vector copy number only in specific Escherichia coli cloning hosts [36]. The copy number of iteron-containing plasmids is normally 1–8 copies/chromosome [37]. Replication initiation from oriV usually requires host DnaA.

Theta-type replication can be either uni- or bidirectional, whereas rolling circle replication is unidirectional [9]. Strand-displacement replication of IncQ plasmids is bidirectional [10]. In most plasmids, theta-type replication is unidirectional (exceptions include the linear Streptomyces plasmid [38]), and there is no replication termination site (exceptions include the plasmid R6K [9,39]).

2.1.2.2. Partition Module. Naturally occurring low copy number plasmids have active segregation mechanisms to avoid plasmid loss upon cell division. These mechanisms are equivalent to the function of the spindle apparatus in a eukaryotic cell [40]. Currently, three types of segregation mechanisms have been proposed [4,5]. Each system consists of a centromere site (often referred to as paras), centromere-binding protein (ParB), and motor protein (ParA). Here, we call a set of the genes and sites for those elements a partition module. A centromere site is generally located directly upstream or downstream of paras genes [41,42]. The segregation mechanism employed by the type I partition system is shown in Fig. 3. In P1 prophage (Table 1), the partition module consists of the paraA-paraB operon and its downstream paras region, which contains multiple ParB binding sites and a host IHF binding site [41]. ParA molecules bound to ATP (ParA+) can bind to DNA non-specifically and thus localize to the nucleoid. The binding of ParB to ParA+ activates ATP hydrolysis by ParA, disrupting the ability of ParA to bind to DNA and releasing it from the nucleoid. Once ParA+ is cleared, the ParB/plasmid complex diffuses through the nucleoid until it makes contact with ParA−. ParB/plasmid complexes in close proximity generate repulsive forces as they clear ParA+ between them. Therefore, replicated plasmid copies are respectively pulled to the opposite ParA−-dense area following the gradient of ParA− (Fig. 3) [5].

repABC family plasmids from Alphaproteobacteria [43] carry a replication module and partition module in the same locus (repABC), and the repABC locus has been used as a vector core for certain types of vectors [44].

2.1.2.3. Toxin-antitoxin (TA) Module. Because plasmids are not tightly connected to the chromosome that carries genes essential for bacterial hosts, cell division can generate plasmid-free cells. If the plasmid is lost upon cell division, the plasmid-free cells, which grow faster than plasmid-containing cells can show an increase in relative population size. This phenomenon can be suppressed by a mechanism called postsegregational cell killing, wherein plasmids produce both stable toxin and unstable antitoxin that counteracts the toxin; plasmid loss results in increased toxin levels in the cells, leading to growth inhibition or cell death of plasmid-free cells [45]. The genetic module responsible for this phenomenon is called the TA module. TA modules can be categorized into six groups according to the mechanism of action [11,46,47]. The first TA system discovered is the hok/sok system of plasmid R1 (Table 1) [45], currently classified as a type I TA module, in which sok encodes an antisense RNA that inhibits the translation of the Mok protein, a regulator for Hok toxin, which generates pores in the cell wall. The hok/sok module of plasmid R1 was applied to improve vector maintenance in the chemostat [48]. The ccdA/ccdB module discovered in plasmid F (Table 1) [49] has also been used in biotechnology. The CcdB toxin inhibits the function of the host DNA gyrase. The ccdB gene has been used as a counter selection marker [50], e.g., in Gateway cloning technology and allele replacements in the chromosome [51,52]. By separating the toxin element and the antitoxin element of a TA module into the chromosome and vector, respectively, StabyCloning technology (Delphi Genetics) enables stable maintenance of a protein-expression vector in the Escherichia coli cell population.

Fig. 1. Modular structure of the plasmid genome. An example of the IncP-1 plasmid genome is shown. The replication module is the only element essential for plasmid maintenance. The other modules assist in plasmid maintenance. IncP-1β and IncP-1α are lineages of IncP-1 plasmids. Horizontal arrows indicate the direction of transcription. Each functional module is regulated by a transcriptional regulator encoded in the module block or global regulators encoded in the central control region [188,189].
2.1.2.4. Multimer Resolution Module. Replicated plasmid copies can recombine into a dimer to multimer via homologous recombination; this negatively affects plasmid partition. Naturally occurring plasmids encode a genetic module to resolve this problem. Small mobilizable plasmids use host-encoded proteins (site-specific recombinases XerC and XerD and accessory proteins PepA and ArgR [53]) for their dimer resolution, and the plasmids carry only a cis-acting resolution site (e.g., cer for ColE1 and its related plasmids, psi for pSC101 [54–56]).

Larger self-transmissible plasmids, e.g., IncP-1 plasmids, carry a host-independent multimer resolution module consisting of a site-specific recombinase (resolvase) gene and a resolution site that also functions as a regulatory region for the resolvase gene [57]. Lack of a resolution module on the plasmid appears to be eventually compensated for by the acquisition of a functionally equivalent cointegrate-resolution system of a Tn3 family transposon, according to observations in experimental evolution [58].

2.1.2.5. DNA Transfer Module and Antirestriction. Conjugative transfer is an important feature of plasmids that enables them to spread genetic information among bacteria (current paradigms for conjugation are summarized in [7]). There are self-transmissible plasmids, mobilizable plasmids, and nonmobilizable or nontransferrable plasmids [59]. The self-transmissible plasmids carry all the gene sets and a cis-acting site (oriT) required for mating pair formation and DNA processing, whereas mobilizable plasmids carry the genes and site only for DNA processing. The Ti plasmid of genus Agrobacterium carries two types of DNA transfer modules: (i) tra/trb operons for DNA transfer between bacteria and (ii) a vir operon for DNA transfer between bacteria and plants [60]. Plasmids from gram-negative bacteria generally use a type IV secretion system for DNA transport, whereas some plasmids from gram-positive bacteria use different DNA transport mechanisms [7,61,62].

Fig. 2. Replication module. (A) Replication module of IncP-1 plasmids. Ssb encodes a single-strand DNA binding protein. trfA encodes the Rep protein. (B) Replication initiation of IncP-1 plasmid. Monomeric TrfA bound to iterons opens base pairs in the DUE, and then host DnaA or TrfA itself recruits DnaB onto the unwound DUE. TrfA molecules bind to single-stranded DNA to assist replicosome assembly on one strand via direct interaction with β-clamp, starting unidirectional replication. Lagging strand synthesis is not shown for simplicity. Illustration follows [8], with minor modifications. (C) Replication inhibition by Rep protein dimer. Paired oriV can dissociate via proteolysis or an increased Rep protein monomer to dimer ratio [32].

Fig. 3. Partition module. (A) par locus of P1 prophage (type I partition system). parA encodes Walker-type ATPase. parB encodes centromere-binding protein. parS is centromere. (B-F) Diffusion-ratchet model of plasmid motion. (B) ParA has two states: ParA*, the ATP-bound form, active in binding DNA; ParA, other forms inactive in binding DNA. ParB binds to parS. For simplicity, only two ParB molecules are shown. (C) ParB binding to ParA* activates ATP hydrolysis by ParA. ParA is released from the nucleoid. (D) ParA slowly exchanges ADP with ATP, then returns to the nucleoid surface. (E) When replicated plasmid copies are present in close proximity, a ParA*-free area is generated between them. Each ParB/plasmid complex diffuses until finding its closest ParA*; thus, their interactions are repulsive. (F) ParB/plasmid complexes are pulled to ParA*-dense areas at opposite ends, following the gradient of ParA*. Illustration follows [5], with modifications.

2.1.2.5. DNA Transfer Module and Antirestriction. Conjugative transfer is an important feature of plasmids that enables them to spread genetic information among bacteria (current paradigms for conjugation are summarized in [71]). There are self-transmissible plasmids, mobilizable plasmids, and nonmobilizable or nontransferrable plasmids [59]. The self-transmissible plasmids carry all the gene sets and a cis-acting site (oriT) required for mating pair formation and DNA processing, whereas mobilizable plasmids carry the genes and site only for DNA processing. The Ti plasmid of genus Agrobacterium carries two types of DNA transfer modules: (i) tra/trb operons for DNA transfer between bacteria and (ii) a vir operon for DNA transfer between bacteria and plants [60]. Plasmids from gram-negative bacteria generally use a type IV secretion system for DNA transport, whereas some plasmids from gram-positive bacteria use different DNA transport mechanisms [7,61,62].
Non-self-transmissible plasmids, including IncQ plasmids represented by RSF1010 (Table 1), can be mobilized by self-transmissible plasmids, e.g., by the IncP-1 plasmid RK2 [63]. oriT has been embedded in some cloning vectors to mobilize the vectors into various hosts for which transformation methods have not been established or are inefficient [64-66].

Plasmid gene content analysis revealed that the complete gene set responsible for self-transmissibility is not necessarily conserved across members of each self-transmissible plasmid group, e.g., IncW and IncP-1 [17]. Interestingly, a gene encoding an antirestriction protein, which blocks the host’s restriction system upon plasmid entry into new hosts, was found to be an element of the plasmid core in IncP-1 and IncW [17]. ArdB, KlCA, ArdA, and ArdC homologs can confer antirestriction against the host’s type I restriction-modification system [67-69]. These antirestriction genes may be important for transfer of synthetic plasmids between different bacterial lineages.

2.2.2. Fitness Cost Imposed by Plasmids

To evaluate the contribution of each functional module to plasmid maintenance, a set of highly unstable broad-host-range plasmid vectors based on the RK2 replicon of the IncP-1 group has been constructed [70]. For example, the functionality of the partition module of a IncU plasmid (Table 1), the chromosome partitioning system of Pseudomonas aeruginosa, and the hipAB TA system of the Paracoccus kordatiae var plasmid have been confirmed using these vectors [70].

2.2.4. Selection Markers

Antibiotics have traditionally been used to select plasmid-containing cells in culture in the laboratory. Mainly for biosafety reasons, various antibiotic marker-free selection approaches have been developed [3,71,72]. Some of the tricks used in such approaches are based on plasmid-derived elements: for example, the RNA I and II of plasmid [23,92,93]. The size distribution of sequenced plasmids available in public databases has been studied. For example, sizes for the 4602 completely sequenced plasmids ranged from 744 bp to 2.58 Mb with a mean value of 80 kb, and the mean value of sizes for mobilizable plasmids was smaller than that for transmissible plasmids [59]. Among the 92 plasmids from the IncF, IncH, IncI, IncN, IncP-1, IncW, A/C, IncL/M, IncP-9, IncQ, IncU, PromA, and Ri/Ti groups used in Suzuki et al. [17], sizes for the non-self-transmissible IncQ plasmids (median size of 8.7 kb) were smallest. Among the self-transmissible plasmids belonging to the six

genes as a model of horizontally acquired genes, Porse et al. [86] demonstrated that physiological interaction of the gene products with hosts imposes a greater cost than nucleotide signals, e.g., G + C content and codon usage. The cause of costs may be relevant to the interactions summarized elsewhere [87] (e.g., disruptive interactions with cellular networks). Currently, it is difficult to predict which interactions negatively affect host fitness and plasmid persistence for an arbitrarily chosen host-plasmid pair. Experimental evolution may help reduce the fitness cost imposed by a synthetic plasmid.

Transcriptome disturbance by a plasmid in a new host is initially high, but will be reduced during fitness cost amelioration [81,82]. Moreover, plasmids encoding H-NS-like stealth protein reduce their fitness cost probably by silencing transcriptional activities of genes in the A + T rich region through the binding of H-NS-like proteins [78,88]. In contrast to smaller or nontransmissible plasmids, larger and transferable plasmids carried multiple NAP genes [89,90]. Three different NAPs encoded on plasmid pCAR1 are involved in plasmid stability and its conjugation in the host cells [91]. Therefore, minimizing unnecessary transcription may be important for minimizing the cost imposed by plasmids.

2.3. Constraints in the Genome

Bioinformatics analysis revealed constraints in plasmids with respect to size, sequence composition (G + C content, oligonucleotide composition, and codon usage), and gene direction. These features may be a result of plasmid-host co-evolution, which can stabilize plasmids in host cell populations. It is important to note that the sequence composition can vary among genes/segments within a plasmid/genome [23,92,93].

2.3.1. Size Constraint

The size distribution of sequenced plasmids available in public databases has been studied. For example, sizes for the 4602 completely sequenced plasmids ranged from 744 bp to 2.58 Mb with a mean value of 80 kb, and the mean value of sizes for mobilizable plasmids was smaller than that for transmissible plasmids [59]. Among the 92 plasmids from the IncF, IncH, IncI, IncN, IncP-1, IncW, A/C, IncL/M, IncP-9, IncQ, IncU, PromA, and Ri/Ti groups used in Suzuki et al. [17], sizes for the non-self-transmissible IncQ plasmids (median size of 8.7 kb) were smallest. Among the self-transmissible plasmids belonging to the six
Inc. groups F, H, I, N, P-1, and W, the median value of sizes (kb) was highest for the IncH group (241 kb), followed by those of the IncF (110 kb), IncI (101 kb), IncP-1 (66 kb), IncN (64 kb), and IncW (39 kb) groups (Fig. 4). Because each plasmid has specific range of genome sizes, it may be important to keep plasmid size in the appropriate range considering the replicon type used in the vector.

Plasmid size may be associated with copy number. For 11 plasmids found in *Bacillus thuringiensis* strain YBT-1520, the plasmid sizes (ranging from 2 to 416 kb) and the copy numbers determined by quantitative polymerase chain reaction (ranging from 1.38 to 172) were negatively correlated [94]. Plasmid F, a member of IncF (median size: 1110 kb), is present at 1 or 2 copies per chromosome, whereas the copy number of RK2, a member of the IncP-1 group (median size: 66 kb), is 3–5 copies/chromosome (in the presence of large replication protein TrfA1), or 1–2 copies/chromosome (without TrfA1) [95]. Plasmid pR28, a member of the IncP-9 group (median size: 83 kb) has a copy number of 1.6–3.7/ chromosome [58]. Copy numbers of the IncQ mobilizable plasmids (median size: 8.7 kb) are 10–16/chromosome [96]. Copy numbers of ColE1-related plasmids are 20–44/chromosome [87,97]. Conlan et al. (2014) determined the copy numbers of plasmids in *Enterobacteriaceae* (3 A/C, 6 IncF, 1 IncH2, 8 IncN, and other plasmids) from the average sequence coverage (depths of PacBio and MiSeq reads) of each plasmid relative to that of the chromosome and showed that copy numbers were 1–3/chromosome [98]. Plasmid copy number estimates can vary, depending on bacterial growth conditions and DNA extraction methods used [87,99]. Therefore, copy number data should be interpreted carefully. To the best of our knowledge, there is no database that catalogs the plasmid copy numbers in various hosts under the same experimental conditions. The elucidation of clear features of plasmid maintenance functions associated with copy number or replicon type requires further investigation.

2.3.2. G + C Content

G + C contents vary widely among bacterial genomes, putatively reflecting a balance among biases generated by mutation and selection [100]. Because bacterial genomes have small regions of noncoding DNA and more protein-coding constraints on first- and second-codon positions than on third-codon positions, most of the variations are due to synonymously variable third-codon positions [101,102]. Growth rate experiments in *Escherichia coli* and *Caulobacter crescentus* showed that decreased generic G + C contents at synonymous sites have negative effects on bacterial fitness when gene expression levels are induced [100,103]. Previous studies have reported that small bacterial genomes tend to exhibit low G + C contents, with some exceptions [104], and that intracellular symbionts, such as plasmids and phages, tend to have lower G + C contents than their hosts [92,105,106]. For the 209 plasmids and their host chromosomes, the G + C contents are highly correlated, and in 164 (78.5%) of cases, plasmids had lower G + C contents than their hosts (Fig. 5). Possible explanations for the lower G + C contents of plasmids than those of hosts include the selection of plasmids that tolerated gene silencing by host H-NS [88,107] and reduced nucleotide synthesis costs [105]. Thus, it may be important that the G + C contents of synthetic plasmids match those of the host chromosomes.

2.3.3. Oligonucleotide Composition

The composition of oligonucleotides, such as di-, tri-, and tetra-nucleotides (also known as k-mers, such as 2-, 3-, and 4-mers), has been studied for the characterization and classification of various organismal genomes [108,109]. Plasmids have oligonucleotide compositions similar to those of their host chromosomes [93,109]. The compositional similarities of plasmids and their hosts suggest that plasmids have acquired hosts’ nucleotide compositions due to amelioration by host-specific mutational biases [110]. Thus, possible plasmid-host pairs are predictable based on the similarity of their oligonucleotide compositions [17]. Earlier studies investigated sequence motifs in the IncP-1 plasmids RK2 [111] and R751 [112] and suggested that some sequence motifs (e.g., tetranucleotide and hexanucleotide palindromic sequences acting as restriction-modification sites) may have been eliminated from plasmids through natural selection. Computational analysis of oligonucleotide compositions has been used to identify novel regulatory DNA sequence motifs [113], some of which may be important for stable plasmid maintenance.

2.3.4. Codon Usage

Synonymous codon usage, tRNA abundance, and ribosome density are critical factors for protein synthesis [114–116]. Codons can affect translation efficiency, protein folding, and mRNA stability [117].

In bacteria such as *Escherichia coli* and *Bacillus subtilis*, highly expressed genes (e.g., those encoding translation elongation factors and ribosomal proteins) tend to preferentially use a subset of synonymous codons that are best recognized by the most abundant tRNA species [118,119]. This is considered evidence of natural selection on synonymous codon usage for translational efficiency and/or accuracy (also called translational selection) [114,120]. Previous studies have indicated that the strength of translational selection on chromosomes varies among bacteria and that fast-growing bacteria with more rRNA and tRNA genes are subjected to strong selection pressure [102]. The strength of translational selection also varies among replicons within the same organism; for example, in *Sinorhizobium meliloti*, codon usage of the chromosome and plasmids pSymB and pSymA reflects their importance for competitive cell growth and expression during the free-living stage of the organism [121].

The codon usage of plasmids is not always similar to that of the host chromosome. Measuring the distance between the codon usages of pairs of *Agrobacterium* plasmids (circular and linear chromosomes and plasmids pAt and pTi) revealed that the distances between chromosomes and plasmids are larger than the distances between the two chromosomes (circular and linear) or the two plasmids (pAt and pTi) [122]. For each pair of three *Agrobacterium* species (*Agrobacterium tumefaciens* C58, *Agrobacterium vitis* S4, and *Agrobacterium radiobacter* K84), codon usages of their plasmids, with varying gene contents, are more similar than codon usages of their chromosomes [123]. It remains unclear whether codon usage influences translation efficiency, protein folding, and mRNA stability.
stable plasmid maintenance in hosts, and the fitness cost imposed by plasmids is still unknown.

2.3.5. Gene Direction

Bioinformatics algorithms based on replication strand biases, such as GC skew, defined as \((C - G)/(C + G)\), have been used to predict replication origin and terminus in bacterial chromosomes and plasmids [124–126]. The degree of GC skew is different between plasmids with and without rolling circle replication and is correlated between plasmids and chromosomes of bacteria, suggesting that replication-related mutation and selection determine the strength of GC skew for replicons within the same host [127]. Previous studies reported that coding sequences (5′ to 3′ orientation) in the bacterial chromosome are preferentially located on the template strands for lagging-strand synthesis (also simply referred to as leading strands [128]), and this codirectional bias of replication and transcription is further enriched in essential and/or highly expressed genes [128–131]. It remains unclear whether gene expressivity and essentiality influence the orientation bias of plasmid genes; however, it may be better to carry important genes on the leading strand of the synthetic plasmids, following the trend in the chromosome.

3. Publicly Available Resources

Comparative sequence analyses of closely related plasmids with different features, such as replication, maintenance, transfer, and host range, can provide hypotheses regarding genetic determinants of these plasmid features. Over the past 10 years, plasmid sequence data have been dramatically increased, and convenient bioinformatics tools have been developed to manage and analyze the data. These resources are briefly described in this section.

3.1. Plasmid Sequence Data

High-throughputDNA sequencing has generated a large amount of plasmid sequences, which can be retrieved from the International Nucleotide Sequence Database Collaboration or INSDC: DDBJ, EMBL-EBI, and NCBI (http://www.ncbi.org). As of 2010, the 1,730 complete plasmid sequences in GenBank were obtained from plasmid-sequencing projects (62%) and microbial genome projects (38%) [132]. In 2015, Shintani et al. [59] used the NCBI database to review the classification of completely sequenced plasmids based on their host taxonomy and features of replication and transfer. Based on the NCBI list of plasmid sequences, downloaded on November 22, 2018 from ftp://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS/plasmids.txt, there are currently 14,309 complete sequences of plasmids, of which 115 are from Eukaryota, 196 from Archaea, and 13,998 from Bacteria (8,538 from Proteobacteria, 3,206 from Firmicutes, 747 from Spirochaetes, and 606 from Actinobacteria).

Because INSDC databases covering all available nucleotide data are not always well curated and structured, secondary databases have been developed. For example, the ACLAME database [http://aclame.ulb.ac.be] [133] has been developed and used to investigate the general features of sequenced plasmids, such as their distribution per host species [134]. Orlekle et al. [135] presented a curated dataset of complete Enterobacteriaceae plasmids compiled from the NCBI database [https://figshare.com/s/18de86dcbb8a47dbba41]. The web servers PLSDB [https://ccb-microbe.cs.uni-saarland.de/plsd/b] [136] and pATLAS (http://www.patalas.site) [137] provide a more comprehensive collection of bacterial plasmids retrieved from the NCBI nucleotide database.

3.2. Bioinformatics Tools

Bioinformatics tools can be used to design synthetic plasmids by searching, assembling, and adjusting key factors, including functional module (genes and cis-element) and genome constraints. Table 3 lists bioinformatics tools for plasmids with their URLs.

Recent studies have developed bioinformatics tools for detecting plasmids from whole genome sequencing data, including cBar [138], PlasmidFinder [139], PLACNET [140], plasmidSPAdes [141], Recycler [142], PlasFlow [143], PlasmidTron [144], PlasmidSeeker [145], and MOB-suite [146]. PLACNETw is a web tool based on PLACNET [147]. These tools can be used for plasmid reconstruction by assembling sequencing reads or de novo assembly graphs (PLACNET/PLACNETw, plasmidSPAdes, Recycler, and PlasmidTron) or for plasmid detection in assembled contigs (PlasmidFinder, cBar, PlasFlow, and MOB-suite) or in unassembled reads (PlasmidSeeker, cBar, PlasFlow, PlasmidTron, and PlasmidSeeker) perform k-mer-based analyses to detect plasmids.

4. Concluding Remarks

Plasmids have been used as primal genetic tools for microbial engineering, particularly for nonmodel organisms. In synthetic biology, there have been attempts to build a vector by assembling functional modules [27,150]. Fortunately, the number of known plasmid sequences has increased dramatically in recent years, which has enabled us to detect core genes and co-evolving gene sets for each plasmid group. Plasmid functional modules identified by experimental or bioinformatics methods can contribute to biological parts/module databases, such as Biobrick [26], SEVA [27], and Clostron [150].

Following the rules of the natural plasmid genome, we can design synthetic plasmids. For example, a set of core genes as well as a selection marker or TA system should be included to prevent generation of plasmid-free cells (Table 2). The G + C content of a plasmid should be similar to (and lower than) that of the host, and highly expressed essential genes should be located on lagging strand templates. We also emphasize that optimization of external settings for the plasmid, for example, type of growth medium and presence or absence of spatial structure in the growth environment, could greatly influence plasmid population dynamics. Although further work is needed, a synthetic biology approach, e.g., de novo synthesis of artificial plasmids followed by...
experimental evaluation of plasmid maintenance, may lead to the construction of stable vectors and improve our understanding of why plasmids are so successful as genetic parasites.

**Competing Interests**

The authors declare no competing interests.

**Authors’ Contributions**

Conceptualization: TO; data analysis: HS; resources; MT; writing – original draft preparation: HY, HS, MS, and TO; writing – review and editing: HY, HS, MS, and TO.

**Acknowledgements**

We thank members of the Institute for Advanced Biosciences at Keio University for useful discussions. We also thank three anonymous reviewers for suggestions for improving this review. This work was supported in part by research funding from Keio University, Yamagata Prefecture, and Tsuruoka City to H.S.T. and by Japan Society for the Promotion of Science (JSPS) KAKENHI (grant numbers 26106001 and 26450090 to T.O., 15H05618 to M.S., and 18K06357 to H.Y.). Computational resources were provided by the Data Integration and Analysis Facility, National Institute for Basic Biology. We would like to thank Editage (www.editage.jp) for English language editing.

**References**

[1] Skippington E, Ragan MA. Lateral genetic transfer and the construction of genetic entities in communities. FEMS Microbiol Rev 2011;35:707–35. https://doi.org/10.1111/j.1574-6976.2010.02601.x.

[2] Wiedenbeck J, Cohan FM. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol Rev 2011;35:657–76. https://doi.org/10.1111/j.1574-6976.2011.00929.x.

[3] Oliveira PH, Mairhofer J. Marker-free plasmids for biotechnological applications – implications and perspectives. Trends Biotechnol 2013;31:539–47. https://doi.org/10.1016/j.tibtech.2013.06.001.

[4] Brooks AC, Hwang LC. Reconstitutions of plasmid partition systems and their mechanisms. Plasmid 2017;91:37–41. https://doi.org/10.1016/j.plasmid.2017.03.004.

[5] Baxter JC. Funnell BE. Plasmid partition mechanisms. Plasmids Biol. Impact Biotechnol. Discov, vol 2 American Society of Microbiology; 2014:135–55. https://doi.org/10.1128/microbiolspec.PLAS-0023-2014.

[6] Fernández-López C, Bravo A, Ruiz-Cruz S, Solano-Collado V, Garson DA, Lorenzo-Díaz F, et al. Mobilizable rolling-circle replicating plasmids from gram-positive bacteria – a low-cost conjugal transfer. Plasmids Biol Impact Biotechnol Discov 2014:2:257–76. https://doi.org/10.1128/microbiolspec.PLAS-0008-2013 American Society of Microbiology.

[7] Cabezón E, Ripoll-Rozada J, Peña A, de la Cruz F, Arechaga I. Towards an integrated genomic signature. J Bacteriol 2010;192:6045–55. https://doi.org/10.1128/jb.00277-10.
Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. Front Microbiol 2015;6:2422. doi:10.3389/fmicb.2015.00242.

Loftie-Eaton W, Yano H, Burleigh S, Simmons RS, Hughes JM, Rogers LM, et al. Analysis of the replicon regions of eleven plasmid pSC101 harbors a recombination protein aids horizontal DNA transmission in bacteria. Sci Transl Med 2014;6:214ra29. doi:10.1126/scitranslmed.3005728.

Eberl L, Kristensen CS, Givskov M, Grohmann E, Gerlitz M, Schwab H. Analysis of the multimer resolution system encoded by the parCBA operon during ColE1 dimer resolution. Mol Microbiol 1996;20:613–22. doi:10.1111/j.1365-2958.1996.tb06953.x.

Caspi R, Helinski DR, Pacek M, Konieczny I. Interactions of DnaA proteins from diverse bacterial species. EMBO J 1992;11:753–6. doi:10.1002/j.1460-2075.1992.tb02445.x.

Meyer R. Replication and conjugative mobilization of broad host-range IncQ plasmids. Plasmid 1994;31:171–81. doi:10.1016/0147-619X(94)90066-4.

Zhou K, Li Z, Liu Y, Chen W, Liao L. CRISPR-Cas9–inducible expression vector: test of various genetic module combinations. BMC Biotechnol 2008;8:2. doi:10.1186/1472-6750-8-2.

Christie PJ, Gordon JE. The art of selective killing: plasmid toxin/antitoxin systems 2014:19–21. doi:10.1186/1472-6750-14-19.

Kobayashi K, Kimbara K, Tanaka Y, Hirota T. Comparative analysis of the replicon regions of eleven plasmids. Plasmid 2009;62:57–66. doi:10.1016/j.plasmid.2008.07.004.

Hiraga S, Sugiyama T, Itoh T. Comparative analysis of the replicon regions of eleven plasmids. Plasmid 2009;62:57–66. doi:10.1016/j.plasmid.2008.07.004.

Bartosik AA, Glabski K, Kulinska A, Lewicka E, Godziszewska J, Markowska A, et al. Caspi R, Helinski DR, Pacek M, Konieczny I. Interactions of DnaA proteins from diverse bacterial species. EMBO J 1992;11:753–6. doi:10.1002/j.1460-2075.1992.tb02445.x.

Vandermeulen G, Marie C, Scherman D, Préat V. New generation of plasmid back selection system: advances in safe and efficient DNA transfer technology. Biotechnol Biofuels 2008;1:3–5. doi:10.1186/1754-6803-1-3.

Bartosik AA, Glabski K, Kulinska A, Lewicka E, Godziszewska J, Markowska A, et al. Construction of a replicative plasmid vector allows efficient expression of recombinant proteins in mycobacteria that requires both type IV and type VII secretion. MBio 2014;5:e01744-13. doi:10.1128/mBio.01744-13.

Sota M, Yano H, M Hughes J, Daughdrill Gw, Abdo Z, Forney Lj, et al. Comparative genomics to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. Sci Transl Med 2014;6:214ra29. doi:10.1126/scitranslmed.3005728.

Bartosik AA, Glabski K, Kulinska A, Lewicka E, Godziszewska J, Markowska A, et al. Caspi R, Helinski DR, Pacek M, Konieczny I. Interactions of DnaA proteins from diverse bacterial species. EMBO J 1992;11:753–6. doi:10.1002/j.1460-2075.1992.tb02445.x.

Sota M, Yano H, M Hughes J, Daughdrill Gw, Abdo Z, Forney Lj, et al. Comparative genomics to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. Sci Transl Med 2014;6:214ra29. doi:10.1126/scitranslmed.3005728.

Bartosik AA, Glabski K, Kulinska A, Lewicka E, Godziszewska J, Markowska A, et al. Caspi R, Helinski DR, Pacek M, Konieczny I. Interactions of DnaA proteins from diverse bacterial species. EMBO J 1992;11:753–6. doi:10.1002/j.1460-2075.1992.tb02445.x.

Sota M, Yano H, M Hughes J, Daughdrill Gw, Abdo Z, Forney Lj, et al. Comparative genomics to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. Sci Transl Med 2014;6:214ra29. doi:10.1126/scitranslmed.3005728.

Bartosik AA, Glabski K, Kulinska A, Lewicka E, Godziszewska J, Markowska A, et al. Caspi R, Helinski DR, Pacek M, Konieczny I. Interactions of DnaA proteins from diverse bacterial species. EMBO J 1992;11:753–6. doi:10.1002/j.1460-2075.1992.tb02445.x.

Sota M, Yano H, M Hughes J, Daughdrill Gw, Abdo Z, Forney Lj, et al. Comparative genomics to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. Sci Transl Med 2014;6:214ra29. doi:10.1126/scitranslmed.3005728.

Bartosik AA, Glabski K, Kulinska A, Lewicka E, Godziszewska J, Markowska A, et al. Caspi R, Helinski DR, Pacek M, Konieczny I. Interactions of DnaA proteins from diverse bacterial species. EMBO J 1992;11:753–6. doi:10.1002/j.1460-2075.1992.tb02445.x.

Sota M, Yano H, M Hughes J, Daughdrill Gw, Abdo Z, Forney Lj, et al. Comparative genomics to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. Sci Transl Med 2014;6:214ra29. doi:10.1126/scitranslmed.3005728.

Bartosik AA, Glabski K, Kulinska A, Lewicka E, Godziszewska J, Markowska A, et al. Caspi R, Helinski DR, Pacek M, Konieczny I. Interactions of DnaA proteins from diverse bacterial species. EMBO J 1992;11:753–6. doi:10.1002/j.1460-2075.1992.tb02445.x.

Sota M, Yano H, M Hughes J, Daughdrill Gw, Abdo Z, Forney Lj, et al. Comparative genomics to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. Sci Transl Med 2014;6:214ra29. doi:10.1126/scitranslmed.3005728.

Bartosik AA, Glabski K, Kulinska A, Lewicka E, Godziszewska J, Markowska A, et al. Caspi R, Helinski DR, Pacek M, Konieczny I. Interactions of DnaA proteins from diverse bacterial species. EMBO J 1992;11:753–6. doi:10.1002/j.1460-2075.1992.tb02445.x.

Sota M, Yano H, M Hughes J, Daughdrill Gw, Abdo Z, Forney Lj, et al. Comparative genomics to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. Sci Transl Med 2014;6:214ra29. doi:10.1126/scitranslmed.3005728.
Salmonella typhi that is temperature sensitive for transfer. Nucleic Acids Res 2000; 28:2177–86.

[160] Sampes F, Furuca N, Tachibana K, Saitou Y, Suzuki T, Mizobuchi K, et al. Complete genome sequence of the incompatibility group 1 plasmid R64. Plasmid 2010;64: 92–103. https://doi.org/10.1016/j.plasmid.2010.05.005.

[161] Carraro N, Poulin D, Burrous V. Replication and active partition of integrative and conjugative elements (ICEs) of the SXT/R391 family: the line between ICEs and conjugative plasmids is getting thinner. PLoS Genet 2015;11:e1005298. https://doi.org/10.1371/journal.pgen.1005298.

[162] Praszkier J, Wei T, Siemens K, Pettard A, J. Comparative analysis of the replication regions of IncR, IncK, and IncP plasmids. J Bacteriol 1991;173:2393–7.

[163] Borrell L, Yang J, Pettard A, Praszkier J. Interaction of initiator proteins with the origin of replication of an IncLM plasmid. Plasmid 2006;56:88–101. https://doi.org/10.1016/j.plasmid.2006.04.002.

[164] Shimada N, Kayama S, Shigemoto N, Hisatsune J, Kuwahara R, Nishio H, et al. Complete nucleotide sequence of pK08-34, an IncM plasmid carrying blhen-14 in Klebsiella oxytoca isolated in Japan. Antimicrob Agents Chemother 2016;60:3156–62. https://doi.org/10.1128/AAC.02507-15.

[165] Carattoli A, Seifert SN, Schwindes S, Perreten V, Endmiani A. Differentiation of IncK and IncM plasmids associated with the spread of clinically relevant antimicrobial resistance. PLoS One 2015;10:e0123063. https://doi.org/10.1371/journal.pone.0123063.

[166] Humphrey B, Thomson NR, Thomas CM, Brooks K, Sanders M, Delsol AA, et al. Construction and characterization of ampli- ted conjugation or protein secretion. Nucleic Acids Res 2014;42:5715–26. https://doi.org/10.1093/nar/gku194.

[167] Gilmour MW, Thomson NR, Sanders M, Parkhill J, Taylor DE. The complete nucleotide sequence and environmental distribution of the cryptic, conjugative, broad-host-range plasmid pIP1062 isolated from bacteria of the wheat rhizosphere. Microbiology 2002;148:1637–53. https://doi.org/10.1099/00221287-87-2-301.

[168] Yanagita K, Maejima Y, Nakata H, Tokuda M, Moriiuchi R, Dohra H, et al. Novel Self-Transmissible and broad-host-range plasmids exogenously captured from anaerobic granules or cow manure. Front Microbiol 2018;9:2602. https://doi.org/10.3389/fmicb.2018.02602.

[169] Papagiannitsis CC, Miriagou V, Giakkoupi P, Tzouvelekis LS, Vatopoulos AC. Characterization of pKP1780, a novel IncR plasmid from the emerging Klebsiella pneumoniae ST147, encoding the VIM-1 metallo-

[170] β-lactamase. J Antimicrob Chemother 1996;38:63–71. https://doi.org/10.1093/jac/66.2.195.

[171] Zhang C, Feng Y, Liu F, Jiang H, Qi Z, Lei M, et al. A phase-like IncEY plasmid carrying the mcr-1 gene in Escherichia coli from a pig farm in China. Antimicrob Agents Chemother 2017;61:e02035–16. https://doi.org/10.1128/AAC.02035-16.

[172] Prokaryotic vectors and vectors from both transposon and plasmid sources for cloning and expression in E.coli. Microbiology 2015;161:97–113. https://doi.org/10.1099/mic.0.09507-0.

[173] Variovi H, Iizukah H. Taxonomic studies on a radio-resistant Serratia marcescens isolated from the environment. Nucleic Acids Res 1992;20:2177–86. https://doi.org/10.1093/nar/19.7.2177.

[174] Yanisch-Perron C, Vieira J, Messing J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 1985;33:103–19.

[175] Zhang C, Feng Y, Liu F, Jiang H, Qi Z, Lei M, et al. A phase-like IncEY plasmid carrying the mcr-1 gene in Escherichia coli from a pig farm in China. Antimicrob Agents Chemother 2017;61:e02035–16. https://doi.org/10.1128/AAC.02035-16.